NASA Contractor Report 172009

National Aeronautics and Space Administration (NASA)/ American Society for Engineering Education (ASEE) Summer Faculty Fellowship Program--1987

Volume 1

William B. Jones, Jr., Editor
Texas A&M University
College Station, Texas

Stanley H. Goldstein, Editor
University Programs Office
Lyndon B. Johnson Space Center
Houston, Texas

(NASA-CR-172009-Vol-1) NATIONAL AERONAUTICS AND SPACE ADMINISTRATION (NASA)/AMERICAN SOCIETY FOR ENGINEERING EDUCATION (ASEE) SUMMER FACULTY FELLOWSHIP PROGRAM, 1987. VOLUME 1 Final Reports (NASA) 326 p

N88-14855
--THRU--
N88-14873
Unclas
G3/85 0116644

Grant NGT-44-001-800

November 1987

National Aeronautics and
Space Administration

Lyndon B. Johnson Space Center
Houston, Texas

PREFACE

The 1987 Johnson Space Center (JSC) National Aeronautics and Space Administration (NASA)/-American Society for Engineering Education (ASEE) Summer Faculty Fellowship Program was conducted by Texas A&M University and JSC. The 10-week program was operated under the auspices of the ASEE. The program at JSC, as well as the programs at other NASA Centers, was funded by the Office of University Affairs, NASA Headquarters, Washington, D.C. The objectives of the program, which began in 1965 at JSC and in 1964 nationally, are

a. to further the professional knowledge of qualified engineering and science faculty members;

b. to stimulate an exchange of ideas between participants and NASA;

c. to enrich and refresh the research and teaching activities of participants' institutions; and

d. to contribute to the research objectives of the NASA Centers.

Each faculty fellow spent 10 weeks at JSC engaged in a research project commensurate with his/her interests and background and worked in collaboration with a NASA/JSC colleague. This document is a compilation of the final reports on the research projects done by the faculty fellows during the summer of 1987. Volume 1 contains sections 1 through 18, and volume 2 contains sections 19 through 35.

CONTENTS

THE DETERMINATION OF NUTRITIONAL REQUIREMENTS FOR SAFE HAVEN FOOD SUPPLY SYSTEM (EMERGENCY/SURVIVAL FOODS)

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1987

Johnson Space Center

Prepared by:	Selina Ahmed, Ph.D.
Academic Rank:	Assistant Professor
University & Department:	Texas Southern University Department of Home Economics Houston, TX 77004
NASA/JSC	
Directorate:	Space & Life Sciences
Division:	Man-Systems Division
Branch:	Flight Systems Branch
JSC Colleague:	Charles T. Bourland, Ph.D.
Date:	August 14, 1987
Contract Number:	NGT44-001-800

ABSTRACT

The Space Station Safe Haven Food System must sustain eight crew members under emergency conditions for 45 days. Sustaining the crew members in an emergency situation following the requirements for minimal volume and maximum shelf life is a challenging part of the Space Station planning. Determination of the nutritional requirements for an emergency/survival ration for Space Station Safe Haven is one of the current research objectives at JSC.

Emergency/Survival Foods are defined as a nutritionally balanced collection of high density foods and beverages selected to provide for the survival of Space Station flight crews in contingency situations.

Safe Haven Foods will be used only in emergency situations. Since storage volume is limited, the foods should be highly concentrated. They will be designed to fulfill one purpose: sustaining the life of the crew during a period of emergency. Advances in food science and technology continue to increase the capability to respond to this challenge.

Recommended dietary composition and special considerations during Safe Haven: A careful study of different research findings regarding starvation and calorie restricted diets indicates that a minimum nutritional need close to RDA is an important factor for sustaining an individual's life in a stressful environment. Fat, protein, and carbohydrates are three energy producing nutrients which play a vital role in the growth and maintenance process of human life. A lower intake of protein can minimize the water intake, but it causes a negative nitrogen balance and a lower performance level. Other macro and micro nutrients are also required for nutritional

interrelationship to metabolize the other three nutrients to their optimum level. Therefore, 2000kcal of Space Station Safe Haven Emergency/Survival Food is recommended on the basis of 10% kcal should come from protein; 60% from CHO; 30% from fat, for 45 days. The various options for longer durations (e.g., 180 to 235 days) are also under investigation. There are some foods also taken under consideration as possible candidates for Space Station Safe Haven Foods.

INTRODUCTION

The Space Station Safe Haven Food System must sustain eight crew members under emergency conditions for 45 days. Sustaining the crew members in an emergency situation following the requirements for minimal volume and maximum shelf life is a challenging part of the Space Station planning. Determination of the nutritional requirements for an emergency/survival ration for Space Station Safe Haven is one of the current objectives at JSC.

Emergency/Survival Foods are defined as a nutritionally balanced collection of high density foods and beverages selected to provide for the survival of Space Station flight crews in contingency situations.

Safe Haven Foods will be used only in emergency situations. Since storage volume is limited, the foods should be highly concentrated. They will be designed to fulfill one purpose: sustaining the life of the crew during a period of emergency. Advances in food science and technology continue to increase the capability to respond to this challenge.

The nutritional criteria will be based on the activity level necessary for routine station operations during both Intravehicular Activities (IVA) and Extravehicular Activities (EVA). Following are the assumptions used in determining the Safe Haven Food System for the Space Station:

A. Two module station with resource nodes
B. An 8-member crew with a planned 90-day resupply cycle
C. A modified food system reduced from a proposed 0.66 cu. ft./man/day to 0.20 cu.ft./man/day
D. The weight (packing density) of all equipment and supplies is calculated to be 22 lbs./cu.ft. The logistics module provides 213 cu.ft. of food stowage volume and the galley provides 165 cu.ft. of food stowage volume for routine station operations
E. The amount of crew time required to maintain the system will not exceed 1 hour/45 days
F. The availability of power in a Safe Haven node shall be provided by other systems

The assumptions made are based on previous space flight experience. The numbers provided in the assumptions may change as new information becomes available. The focus of this paper is on the nutritional criteria. (15)

In the selection of foods for the Safe Haven System, consideration needs to be given to their:

A. Functions in the body
B. Nutritional contribution
C. Relationship with the normal daily menus of the Food Supply and Service System (FSSS)

PURPOSE OF THE STUDY

The purpose of the study was to evaluate potential food types and recommend candidate methods for meeting the nutritional requirements for the Space Station Safe Haven Food System.

OBJECTIVES

1. To determine the overall nutritional requirements for Safe Haven food supply systems
2. To recommend an emergency ration providing the minimum number of calories/man/day required to prevent weight loss
3. To establish a knowledge base of the specific nutrients required and their metabolic effect on the health of Space Station crew members in terms of stress and performance.
4. To determine the minimum water requirements for Safe Haven
5. To determine the minimum Safe Haven Food shelf life requirements
6. To recommend potential Safe Haven foods

REVIEW OF THE LITERATURE

C. F. Consolazio and his associates (1971) conducted research in Metabolic Imbalances and Body Hypohydration During Food Deprivation (10 days). In the initial study of 6 men who fasted completely for 10 days, significant metabolic stresses developed which could have eventually led to serious abnormalities. These observations included a great amount of body dehydration which resulted in substantial body weight loss in the test subjects, coupled with large nitrogen and mineral losses, and a marked ketosis. (10)

Three studies on complete starvation and restriction of caloric intake, each lasting 10 days, will be discussed. In the first study, 6 healthy adult males between the ages of 21 and 52 years were placed on a starvation diet for 10 days. Water was available ad libitum. (13)

The major problems encountered during 10 days of complete starvation were:

A. Large body weight losses - averaged 7.27 kg or 9.5% of the total body weight
B. Highly negative water balances resulting in great body dehydration
C. Negative nitrogen balances indicating that excessive body protein was being catabolized
D. A marked ketosis
E. Large mineral losses
F. Abnormal EKG's in all subjects during the experimental phase

In the second study, the subjects were 8 young male volunteers who engaged in an intensive physical training program prior to the onset of the study. The study consisted of 3 phases: a control period of 8 days, followed by a 10-day period of caloric restriction, and then an 8-day rehabilitation

phase. Group I received no mineral supplementation, while Group II was given mineral supplements. No vitamin supplements were given to any of the subjects during the caloric restriction period. The diet used during the restriction period contained 420 cal/day of carbohydrate. Energy expenditure was maintained at the 3200 cal/day level. The second study was designed to minimize protein catabolism, decrease electrolyte excretion, eliminate ketosis, maintain the water balance by feeding a small quantity of carbohydrates (100 gm), and to observe the effect of mineral supplements on these metabolic factors. The study showed that mineral supplementation under the experimental conditions did not affect the nitrogen balance.

The third study, a 500 cal/day diet, showed that 500 calories/day are inadequate for short term performance. Although they spared water, the protein catabolism was still a major problem. (13)

According to H. J. Krzywicki, et.al. (1972) in their 420 Kcal study, weight loss was 15 percent greater when minerals were totally excluded. The losses were less severe than the 9.5 percent loss observed during the starvation study conducted by Krzywicki in 1968. The addition of 40 gm of protein to the diet resulted in losses of 7.6 percent in Group I and 6.5 percent in Group II, which were not dissimilar from the 8.1 and 5.9 percent losses observed during the 420 Kcal study. (19, 20)

The weight losses in Group I closely approximated values reported in the Brozek study (7.7 percent loss after 12 days on a 500 Kcal carbohydrate intake). Observations made by Grande, et.al., show 4.5 percent losses in body weight when male test subjects were fed a 1000 Kcal/day carbohydrate diet for 16 days. They also had negative water balances in spite of a 4.5 gm/day NaCl intake. Therefore, neither the 1000 Kcal, 40 gm of protein, nor mineral supplements prevented water loss. (6, 16)

The weight loss was calculated to be 65 percent fat in Group I and 59 percent fat in Group II in the study conducted by Krzywicki (1979). This was markedly different from the 22-24 percent loss attributable to fat when a diet of 420 Kcal of carbohydrates was used. Brozek reported that fat loss constituted 40 percent (190 gm/day) of the total body weight lost. Group I lost 360 gm/day, while Group II lost 276 gm/day of fat. This suggests that fat stores were mobilized and utilized to a greater extent when protein was added to the diet.

In the study by Krzywicki, protein values are in agreement with the findings of Brozek who reported that 9 percent of the body weight loss was protein. In the Krzywicki study, 390 gm of dry protein constituted 7 percent of the body weight lost in Group I and 8.4 percent in Group II, in view of the fact that the protein estimate was calculated as only 20.2 percent of the fat free mass. Water losses were reduced, but some dehydration occurs with caloric restriction. In conclusion, during the 500 Kcal/day diet comprised of 85 gm of carbohydrates and 40 gm of protein, the weight loss was comprised of 59 to 65 percent fat, 8.4 percent of dry protein, and 25 to 28 percent water. (6, 21)

Johnson, et.al. (1971), in his Metabolic Aspects of Calorie Restriction Study, showed that negative nitrogen balances averaged 7.1 and 6.5 gm/man/day, and no improvement in nitrogen balances was observed. The protein loss accounted for approximately one-fourth of the weight loss in the experimental group. (17)

The U.S. Army Research Institute of Environmental Medicine, located in Natick, MA, conducted a study relating nutritional status with the physical

and mental performance of special operations troops consuming the Ration, Light-Weight (RLW), or the Meal, Ready-to-Eat (MRE), military field ration. The study was conducted during a 30-day field training exercise. The MRE is designed to provide adequate nutrition under operational conditions and provides 3900 Kcal/day (3 packages, 1300 Kcal/package). The RLW-30 was designed for special operations that do not permit resupply over an extended period of time (30 days). A 30-day field test of the RLW-30 ration showed that the troops were adequately sustained with a minimal weight loss. All participants stayed within the acceptable weight loss of 6.3% of their original body weight. The MRE was tested at the same time as a control group and was found to have sustained the group to within 2.2% of their original body weight. The significant difference between the weight losses in the study was that the MRE group lost only body fat, whereas the RLW-30 group lost both lean body mass as well as body fat. (1)

It has been shown that many abnormalities are associated with long term starvation. Canadian prisoners of war (POWs), who had been on restricted diet in camps in Singapore and Hong Kong during World War II, were subjects of a study conducted more than 10 years after they were released. Some of the most frequent symptoms observed in the study group included: fatigue with low levels of effort, profuse sweating for no apparent reason, numbness and cramps in the calf muscles, loss of ambition, poor vision, edema, dyspnea with even the slightest exertion, depression, anorexia, nausea, restlessness, irritability, and insomnia. (7)

DISCUSSION

The intended purpose of the Safe Haven Food System is to allow flight crew members to continue to live and function in space in an emergency situation. During an emergency, the regular food supply may be inaccessible by loss of a node or module or rendered unusable by accident, lack of power, etc., necessitating the use of a contingency food system. The current provisions for Space Station emergency management are Safe Haven and the Health Maintenance Facility (HMF), which together include: (15)

Water
Food
Food Preparation Equipment
Utensils
Wipes
Waste Management Provisions
Trash Management Provisions
Personal Hygiene Provisions
Towels
Clothing
Sleep Restraints
Health Maintenance Provisions
Lighting
Exercise Equipment
Communication Equipment
Tools

First Aid Kit

Possible Space Station emergencies may be caused by:

Fire
Collision with Orbital Debris
Loss of Temperature Control
Loss of Pressurization
Contamination
Loss of Attitude Control
Explosion
Meteoroid Penetration
Grazing Collision
Illness/Injury
Depletion of Consummables
Inadvertent Operation

Should such an emergency occur, a decision would have to be made either to allow the crew to continue to live on the Station for up to 45 days to wait for the next launch of the Shuttle, or to return the crew to Earth in a Crew Emergency Return Vehicle (CERV). The first option would require the crew to maintain, or regain and maintain control of the Station. If all systems fail, the decision would be to bring the crew members home safely. Either decision will require a special food system to maintain the health of the crew members until a rescue can be completed or the emergency situation is corrected.

The CERV will be a dedicated rescue vehicle which will be attached to a node. During a rescue using the CERV, the maximum stay in the vehicle will be one day in orbit and one day on the ground. The CERV will have sufficient food and water stowage for a crew of 8 for the two-day period. A menu needing a minimum of preparation and no preparation equipment will be required in the small quarters of the CERV. During a rescue mission, use of survival type food would be a more logical approach than use of a nutritionally balanced diet. The crew will have no special physical activities which would require the normal daily food supply of 3000 Kcal. A high density food bar containing 500 to 900 Kcal should be sufficient to meet the temporary needs.

The rationale of this recommendation is based on:

A. Weight and volume limitations
B. Lack of sufficient power
C. Space available for food preparation
D. Quality and quantity of food required

The research findings from the various sources indicate that, during complete starvation, body dehydration results in conjunction with substantial body weight losses, large nitrogen and mineral losses and marked ketosis. These findings are not unusual since both the body fat and protein stores must be utilized as energy sources. The maintenance of normal blood carbohydrate levels requires a known quantity of protein breakdown. A low antiketogenic diet and adequate mineral supplements could prevent the marked ketosis, minimize protein catabolism, maintain fluid balance, and decrease

the electrolyte excretion. The large losses of total nitrogen in urine during caloric restriction are still indicative of catabolism of body protein for gluconeogenesis and energy. Studies have indicated that the body has labile stores of protein reserves which are readily lost during adaptation to low protein or low caloric diets. (8, 9, 11, 18)

The recommended Space Station menu developed by the Food Supply and Service System study team conforms to the current Government nutritional guidelines, and contains as a minimum:

A. Four servings of fruits and vegetables
B. Two servings of meat, poultry, fish, or vegetable protein equivalent
C. Four servings of grain based (cereal) foods, with an emphasis on whole grain foods where possible
D. Two servings of milk or its calcium equivalent

(see TABLE I - RECOMMENDED DAILY DIETARY ALLOWANCES OF NUTRIENTS)

TABLE I - RECOMMENDED DAILY DIETARY ALLOWANCES OF NUTRIENTS

NUTRIENT	DAILY ALLOWANCE (1)	
	MALES	FEMALES
Kilocalories, kcal	2300-3100 (2)	1600-2400 (2)
Protein (3), g/kg	0.8	0.8
Vitamins		
Thiamin, mg	1.4	1.0
Riboflavin, mg	1.6	1.6
B6, mg	2.2	2.0
B12, µg	3.0	3.0
Folacin, µg	400	400
Niacin, mg	18	13
C, mg	60	60
A, µg, (as retinal equivalents)	1000	800
D, µg (as cholecalciferol)	5	5
E, mg (as tocopherol equivalents)	10	8
Minerals		
Iron, mg	10	18
Calcium, mg	800	800
Phosphorus, mg	800	800
Magnesium, mg	350	300
Zinc, mg	15	15
Iodine, µg	150	150
ESTIMATED SAFE AND ADEQUATE INTAKES FOR ADULTS		
Vitamins		
Pantothenic acid, mg	4-7	
Biotin, µg	100-200	
K, µg	70-140	
Electrolytes		
Sodium, mg	1100-3300	
Potassium, mg	1875-5625	
Chloride, mg	1700-5100	
Minerals		
Manganese, mg	2.5-5.0	
Selenium, µg	50-200	
Molybdenum, µg	150-500	
Copper, mg	2.0-3.0	
Chromium, µg	50-200	
Fluoride, mg	1.5-4.0	

Levels of nutrients are based on the NAS, NRC, and the Committee on Dietary Allowances.

(1) Allowances are for adults, age 23-50

(2) Kilocalorie allowance may vary depending on body size, weight, and work load.

(3) Includes a balanced (nonlimiting) level of the nine essential amino acids, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine.

Current research findings of the nutritional requirements of man in space, and ground based studies on the impact of restricted diets on the survival of human life indicate that there is a crucial need to recommend a nutritional guideline for Safe Haven foods for longer duration use. The Safe Haven menu should follow as closely as possible the recommended daily allowances (RDA). (15,22)

NUTRITIONAL REQUIREMENTS

Recommended dietary composition and special considerations during Safe Haven:

1. The large losses of total nitrogen in urine during caloric restriction are still indicative of catabolism of body protein for gluconeogenesis and energy. Therefore, a minimum of 10 percent protein (40 gm) should be included in the Safe Haven Food System. A limit of up to 10 percent protein can minimize the water intake required.
2. During complete starvation, significant metabolic stress can develop which eventually may lead to serious abnormalities such as: body dehydration, nitrogen loss, mineral loss or a marked ketosis. Hence, the Safe Haven Food System should be comprised of at least 60 percent (300 gm) carbohydrates which may prevent ketosis and other abnormalities. The addition of 40 gm of protein to the calorie restricted diet may also increase the utilization of fat.
3. A body weight loss is normally attributed to a calorie restricted diet. Brozek reported loss of fat contributed 40% of the total body weight loss in his low calorie carbohydrate diet. Therefore, at least 30 percent (66 gm) fat should be included in the Safe Haven Food System. (6)
4. The moisture content of each food item deserves consideration as a natural source of water to maintain body fluid levels.
5. Preparation time is an important criteria in the selection of food items for Safe Haven.

TABLE II - PROPOSED FOOD MIX & FOOD TYPES

FOOD MIX	FOOD TYPE	PERCENT OF FOOD ITEMS	PERCENT OF BEVERAGE ITEMS
50% MRE	THERMOSTABILIZED	50-60	---
50% WITH MINIMUM PREPARATION	INTERMEDIATE MOISTURE/NATURAL FORM	20-30	---
	REHYDRATABLE	15-20	100

The appearance, odor, flavor, and familiarity of Safe Haven Food for Space Station should be comprised of typical Earth-like characteristics, but not limited to it under the special circumstances. It must be of optimum utility and also have choice in menu selection on board the Space Station. A minimum of 7 and possibly of a 12-day menu cycle is also suggested criteria for Safe Haven.

According tothe Manned Module Configuration, there will be Hab and Lab in Space Station. In the Lab section of Space Station, a possible suggestion is to keep a food warmer like Shuttle's to provide a facility for minimum food preparation. Therefore, determination of percentage of power utilization during food preparation at minimum rate needs to be studied. An easy access to food, water, and accessory hardware items should be facilitated by the system. A specific routine inventory should be required for leakage, bulges, tears, and other anomalies in sufficient time for each resupply mission. (15)

The characteristics and discussions about possible Space Station Safe Haven Foods (Emergency/Survival) are as follows:

TABLE III - MAJOR CHARACTERISTICS OF CANDIDATES

RATION	PROPERTY OF FOOD CONTENTS
Meal ready to eat (MRE)	• Thermostabilized • Contains Water • Palatable • Ready to eat • 84-month maximum storage at 70 degrees F
Ration lightweight 30 days (RLW-30)	• Dehydrated • Low volume • No water - needs water source • 36-month storage at 100 degrees F
Nutritional sustainment module (NSM)	• Dehydrated • Low volume • No water • High calorie value
Shuttle foods (emergency survival ration)	• Thermostabilized • Dehydrated

MRE (Meal Ready-To-Eat)

The Meal Ready-To-Eat provides individual meals containing food components that are ready to eat under conditions precluding preparation, except reconstitution of beverages. The MRE was developed for use in a combat zone. The MRE has 12 different menus which could serve a 12-day menu cycle of Space Station Safe Haven. After consideration and evaluation

of its quality characteristics, MRE foods could be possible candidates for Space Station Safe Haven Foods. The minimum limited shelf life for all emergency food types under Space Station stowage conditions is 3 years, but a recommendation of 7 years or more is possible. Therefore, more ground based studies under recommended optimum conditions are advisable. (1, 15)

RLW-30

A 2000 Kcal light weight ration (RLW-30) was tested as the sole source of food for 30 continuous days during a Special Forces field training exercise. Medical examinations did not reveal any serious medical problems, and there was no evidence of direct ill effects from the ration. Some members of the RLW-30 group noted trace urinary protein loss. Nutrient intake was adequate to meet military dietary allowances. The macro nutrients are intentionally reduced in the ration to meet size/weight constraints and reduce the water burden of the ration. Studies have not yet been completed on the limited life of the RLW-30. Therefore, a ground based study on longer shelf life beyond 7 years will be recommended. (1, 15)

Nutritional Sustainment Module

As currently specified, the Nutritional Sustainment Module (NSM) represents an entirely new concept for achieving compactness, convenience, and nutritional reliability in a combat ration. This has been developed to meet the challenge of the 21st century's stressful war environments. It is of the highest possible caloric density, approaching 7.1 Kcal/cc. A dismounted soldier will be able to easily carry a 3-5 day food supply meeting the 3600 Kcal daily requirements. Another objective of the NSM is optimum utility, consummable on the move without any preparation, and modular in design. For example, a 900 Kcal module would be able to combine with others to form rations of 1800, 2700, 3600, 4500 Kcal to meet specific combat scenario requirements.

A field test was conducted on the NSM 900 Kcal module. The positive characteristics of these rations are that it is compact, light, easy to carry, and easy to open for consumption. A module demonstrator has been developed with the following characteristics: 900 Kcal, 160 cc, 150 gm, 5.6 Kcal/cc, and 6.0 Kcal/gm. This demonstrator module exceeds the design criteria for caloric density, and represents an 86% reduction in the cube and a 70% reduction in weight over the current MRE.

The available shelf life of NSM as of today is 3 years. With the advancement of new technology in food science, a shelf life of NSM beyond 7 years is possible for subsisting in the battlefield of the future. (2)

Shuttle Foods

A few selected high density Shuttle foods are also future candidates for Space Station Safe Haven food service system. (15)

The Food Development Lab of the Johnson Space Center at Houston conducted the various studies regarding stowage volumes and weight for potential Safe Haven Foods for a 45-day supply. Table IV indicates the results of the study. (3, 4, 5)

TABLE IV - STOWAGE VOLUMES AND WEIGHTS FOR POTENTIAL SAFE HAVEN FOODS (45-DAY SUPPLY, CREW OF 8)

FOOD SYSTEM	FOOD WEIGHT KG (LB)	MAKUP WATER WEIGHT KG (LB)	TOTAL WEIGHT KG (LB)	FOOD VOLUME CU.M. (CU.FT.)	WATER VOLUME (a) CU.M. (CU.FT.)	TOTAL VOLUME CU.M. (CU.FT.)	KCAL/MAN/ DAY
MRE (2 pkg/man/ day)	340 (750)	836 (1843)	1176 (2593)	.60 (21)	.84 (29.6)	1.43 (50.6)	2600
RLW-30 (1-1/4 pkg/ man/day	204 (450)	998 (2200)	1202 (2650)	.34 (12)	1.0 (35.3)	1.34 (47.3)	2460
Standard Shuttle	571 (1260)	697 (1537)	1269 (2797)	1.70 (60)	.70 (24.7)	2.40 (84.7)	2750
Emergency Shuttle	317 (700)	937 (2066)	1255 (2766)	.71 (25)	.94 (33.2)	1.65 (58.2)	2670

(a) This volume does not include the containers for the water or any dispensing system.

Safe Haven Foods will be used only when there will be no Crew Emergency Rescue Vehicle (CERV) available and will allow 8 crewmembers to live for 45 days. A Nutritional Sustainment Module (NSM) can be used for CERV because of its preferable characteristics: compact, light, easy to open for consumption, and longer shelf life. In this paper, a 45-day mission was emphasized to establish the nutritional requirements for the Space Station Safe Haven food supply. An optional plan is under consideration for a more extended mission. Option I is the 45-day survival mission. Option II will be discussed as follows:

OPTION II

ASSUMPTIONS:

--Eight crew members
--Ninety-day duration - possible duration up to 180 to 235 days
--Two crew members will be assigned to perform extravehicular activities (EVA)
--All crew members will have IVA
--Induced reduction metabolic rate for crew members
--Available recycling water
--Stowage unknown
--High density foods will be used; Nutritional Sustainment Module (NSM) is a possible candidate
--Shelf life of food should be beyond 7 years

In Option II, ground-based research needs to be done to observe the cause and effect of an induced reduction in the metabolic rate. A mission of 180 to 235 days would be possible using different methods of reducing metabolic rates in human beings. This may involve the use of new developments in medical science. There are several methods which can be used for reduction in metabolic rate, i.e., decrease activity, sleep medication on a chronic basis, meditation, and others. The results of the ground-based studies will help to understand the percentage of reduction in the metabolic rate and will aid in planning and establishing nutritional needs for the future crew members of Space Station in similar situations.

In various research studies, the results indicate that starvation for 10 days can cause large body weight loss, high negative water balance resulting in body dehydration, negative nitrogen balance, a marked ketosis and abnormal EKG during the experimental phase; human performance remains remarkably good even under suboptimal nutrient intake for periods up to 24 days. (12, 14, 18) Therefore, if the crew members of Space Station plan to survive in an emergency situation for 180 to 235 days with minimum nutrients, an above mentioned alternative medical and nutritional plan needs to be under consideration. A temporary induced reduction in metabolic rate will minimize several problems in the contingency plan and will help to establish the nutritional needs for the crew members for longer durations.

SUMMARY

The concept of Space Station Safe Haven is to sustain 8 crew members under emergency conditions for 45 days. After careful study of different research findings in respect to starvation and calorie restricted diet, a minimum nutritional need close to the RDA is an important factor for sustaining an individual's life in a stressful environment. Fat, protein, and carbohydrates are three energy producing nutrients that play a vital role in the growth and maintenance process of human life. A lower intake of protein can minimize the water intake, but it shows significant effects of negative nitrogen balance and the lower performance level of the individual. Other macro and micro nutrients are also required for a nutritional interrelationship by allowing the first three nutrients to work in the body up to their optimum level. Therefore, 2000 Kcal Space Station Safe Haven Emergency/Survival food is recommended on the basis that 10% Kcal should come from protein, 60% from CHO, and 30% from fat, for 45 days.

FUTURE RECOMMENDATIONS

1. Adaptation training for different calorie restricted diets, e.g., 300 Kcal up to 900 Kcal.
2. Psychological testing for Space Station crew members concerning calorie restricted diet, as well as psychological impact from the absence of normal daily diet.
3. Physical training (i.e., exposure) regarding the stress and fatigue associated with significantly calorie restricted diet.
4. Study the crew members on the basis of mass and sex, and their impact on consumption of daily food in comparison with standard male and female size.

REFERENCES

1. Askew, E. W., Munro, I., Sharp, M. A., Siegel, Popper, Rose, et.al., Nutritional Status and Physical and Mental Performance of Special Operations Soldiers Consuming the Ration, Lightweight, or the Meal-Ready-To-Eat Military Field Ration During a 30-Day Field Training Exercise. Report No. T7-87 U.S. Army Research Institute of Environmental Medicine, Natick, MA, March 1987.

2. Briggs, J., Duane, C. P., Graham, M., Risvik, E., Cardello, A., Barrett, A., Taub, I. A. A Calorically Dense Ration for 21st Century. U.S. Army Natick Research, Development & Engineering Center, Natick, MA.

3. Bourland, C. T. Heating Food in Space. Memo, 1985.

4. Bourland, C. T., Fohey, M. F., Rapp, R. M., and Sauer, R. L. Space Shuttle Food Processing and Packaging. Journal of Food Protection, 44:313-319, 1981.

5. Bourland, C. T. Shuttle Pre-Launch and Post-Landing Food Service Operational Procedures. JSC-14887, NASA, 1984.

6. Brozek, J., Grande, F., Taylor, H. L., Anderson, J. T., Buskirk, E. R., and Keys, A. Changes in Body Weight and Dimensions in Men Performing Work on a Low Calorie Carbohydrate Diet. Journal of Applied Physiology, 10:412, 1957.

7. Coke, L. R. Late Effects of Starvation. Medical Services Journal of Canada, 17:313, 1961.

8. Consolazio, C. F., Matoush, L. O., Johnson, H. L., Nelson, R. A., and Krzywicki, H. J. Metabolic Aspects of Acute Starvation in Normal Humans (10 Days). American Journal of Clinical Nutrition, 20:672, 1967.

9. Consolazio, C. F., Nelson, R. A., Johnson, H. L., Matoush, L. O., Krzywicki, H. J., and Isaac, G. J. Metabolic Aspects of Acute Starvation in Normal Humans: Performance and Cardiovascular Evaluation. Amercian Journal of Clinical Nutrition, 20:684, 1967.

10. Consolazio, C. F., Matoush, L. O., Johnson, H. L., Krzywicki, H. J., Isaac, G. J., and Witt, N. F. Metabolic Aspects of Calorie Restriction: Hypohydration Effects of Body Weight and Blood Parameters. American Journal of Clinical Nutrition, 21:793, 1968.

11. Consolazio, C. F., Matoush, L. O., Johnson, H. L., Krzywicki, H. J., Isaac, G. J., and Witt, N. F. Metabolic Aspects of Caloric Restriction: Nitrogen and Mineral Balances and Vitamin Excretion. American Journal of Clinical Nutrition, 21:803, 1968.

12. Consolazio, C. F., Johnson, R. E., Pecora, L. L. Physiological Measurements of Metabolic Functions in Man. 1242. McGraw-Hill Book Company, Inc., 1963, p. 340.

13. Consolazio, C. F., Johnson, H. L., Krzywicki, H. J. Metabolic Imbalances and Body Hypohydration During Food Deprivation (10 Days). Bioenergetic Division, U.S. Army Medical Research and Nutrition Laboratory, Fitzsimons General Hospital, Denver, CO, March 1971.

14. Consolazio, C. F. Nutrition and Physical Performance. Edited by R. E. Johnson. Pergamon Press, NY, 1983.

15. FSSS Study Team. Space Station Food Supply and Service System Specifications Document, SS-F-0003, August 1987.

16. Grande, F., Taylor, H. L., Anderson, J. T., Buskirk, E. R., and Keys, A. Water Exchange in Men on a Restricted Water Intake and Low Calorie Carbohydrate Diet Accompanied by Physical Work. Journal of Applied Physiology, 12:202, 1958.

17. Johnson, H. L., Consolazio, C. F., Krzywicki, H. J., and Witt, N. F. Matabolic Aspects of Calorie Restriction: Nutrient Balances with 500 Calorie Intakes. American Journal of Clinical Nutrition, 24:913, 1971.

18. Keys, A. H., Taylor, L., and Brozek, J. Nutrition and Performance Capacity. Research and Development Command, Quartermaster Corps, U.S. Army, Report #7, 1956.

19. Krzywicki, H. J., Consolazio, C. F., Matoush, L. O., and Johnson, H. L. Metabolic Aspects of Acute Starvation: Body Composition Changes. American Journal of Clinical Nutrition, 21:87, 1968.

20. Krzywicki, H. J., Consolazio, C. F., Johnson, H. L., and Witt, N. F. Metabolic Aspects of Caloric Restriction (420 Calories): Body Composition Changes. American Journal of Clinical Nutrition, 25:67, 1972.

21. Krzywicki, H. J., Consolazio, C. F., Johnson, H. L., Witt, N. F. Metabolic Aspects of Caloric Restriction (500 Calories): Body Composition Changes. Letterman Army Institute of Research, San Francisco, CA, August 1979.

22. Stadler, C. R., Bourland, C. T., Rapp, R. M., and Sauer, R. L. Food System for Space Shuttle Columbia. Journal of the American Dietetic Association, 80 (2):108-114. 1982.

LIFE SCIENCES FLIGHT EXPERIMENTS MICROCOMPUTER

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1987

Johnson Space Center

Prepared by:	Peter N. Bartram, Ph.D.
Academic Rank:	Associate Professor
University and Department:	Norwich University Computer Science Engineering Division of Engineering Northfield, Vermont 05663
NASA/JSC	
Directorate:	Space and Life Sciences
Division:	Life Sciences Project
Branch:	Project Engineering
JSC Colleague:	James S. Evans
Date:	August 6, 1987
Contract Number:	NGT 44-001-800

ABSTRACT

A promising microcomputer configuration for the Spacelab Life Sciences Laboratory Equipment inventory consists of multiple processors. One processor is reserved for the principal investigator's use, with additional processors dedicated to real-time input and output operations. A simple form of such a configuration, with a processor board for analog-to-digital conversion and another processor board for digital-to-analog conversion, was studied. The system used digital parallel data lines between the boards, operating independently of the system bus. Good performance of individual components was demonstrated: the analog-to-digital converter operated at over 10000 samples per second. The combination of the data transfer between boards with the input or output functions on each board slowed performance, with a maximum throughput of 2800 to 2900 analog samples per second. Any of several techniques, such as use of the system bus for data transfer or the addition of direct memory access hardware to the processor boards, should give significantly improved performance.

INTRODUCTION

The Life Sciences Laboratory Equipment (LSLE) inventory for the Spacelab includes a microcomputer for high-speed acquisition of Life Sciences experimental data (both analog and digital). This computer provides the vital link between Spacelab experiments aboard Shuttle missions and the Science Monitoring Area in the Life Sciences Project Divison building at the Johnson Space Center. The LSLE microcomputer sends the real-time data, in digital form, properly formatted, to the High Rate Multiplexor (HRM), which in turn transmits the data to Earth via communication (TDRSS) satellite. The current LSLE microcomputer is based upon compatibility with the PDP-8, and thus no longer meets the goal of being compatible with systems familiar to principal investigators in the life sciences.

Requirements for a replacement LSLE microcomputer include providing all of the data acquisition, HRM downlink, and experiment control functions of the current machine. Additional capabilities needed include better and more extensive displays in the Spacelab, faster data sampling rates, extensive data storage for use with mid-deck experiments (no HRM downlink available for missions not carrying the Spacelab), greater speed, more memory, compatibility with modern computer systems (such as the IBM PC) familiar to principal investigators, and the ability to be programmed using higher level languages. The design should be based upon commercially available boards and a recognized backplane bus to connect them. Custom hardware should be avoided (HRM interface excepted).

Of the possible system architectures, a multiple processor (CPU) approach seems to offer the greatest promise. By use of separate CPU's for data acquisition and control (for example, analog-to-digital conversion or output to the HRM), a "master" CPU board could be reserved almost entirely for the principal investigator's use. Expansion would have minimal impact on the original parts of the system, as additional processors would handle the additional data acquisition load. This modularization would facilitate software and hardware maintenance. It would allow on-going replacement of obsolete components at the board level, without significantly disturbing other parts of the system.

EQUIPMENT USED

The Intel 8088 processor and boards using this processor were chosen for compatibility with familiar microcomputers (such as the IBM PC) and because of the availability of a wide range of software, including programming languages. The STD bus was selected because of the significant number of processor and other STD bus boards available from several vendors. Boards meeting the STD bus standards are also of relatively small size. To demonstrate the concept of the multiple processor configuration, a simple case was studied. This system consists of a master processor board (with memory), reserved for the principal investigator's use, and additional processor boards (with separate memory) for use with peripheral devices. Each peripheral board also contains an interface for real-time modules attached to the board by means of the Intel iSBX connector. Data transfer from memory on the peripheral boards to other boards can be accomplished using the STD bus (but only by the master processor board, as the peripheral boards cannot control the STD bus), or independently of the master board using a separate external direct parallel interface. One board is equipped with a multi-channel analog-to-digital converter, the other with a digital-to-analog converter. The digital-to-analog converter served as a representative output device, chosen here as a simpler and more readily available output device compared the HRM, while involving the same process of accepting data from another board in the system. In the final system, many additional boards will be needed, including digital-to-analog converters and a custom HRM interface board, as well as digital input and output interfaces. Each would be contained on its own dedicated processor board on the STD bus.

For the small scale demonstration system studied, Table 1 lists the equipment used. The master CPU, not listed in the table, was not actually used, as it was not needed for the peripheral board to peripheral board data transfer (STD bus lines were not used for this). Rather, each peripheral processor board comes with the Intel 8256 parallel input/output chip, which provides a means of independent data transfer, complete with interrupts and "handshake" control signals.

TABLE 1

EQUIPMENT USED

HARDWARE

Ziatech 8862 Card Cage and Power Supply (STD bus)
Ziatech 8830 Intelligent I/O Control Processor Board for STD bus (2 used)
Data Translation DTX311 Analog-to-Digital Converter module
Data Translation DTX328 Digital-to-Analog Converter module
personal computers (2 used) for downloading programs to the Ziatech 8830 boards and debugging them on-board.
oscilloscopes for displaying analog output and parallel transfer handshake signals.
function generator for analog input signals.

SOFTWARE

Ziatech 8830 software (and ROM's) to provide program loading and debugging on-board, using the personal computer and the 8830 serial port.
Microsoft C Language, version 4.00
Microsoft Macro Assembler, version 4.00
Microsoft Linker (loader), version 3.51
Mircosoft MS-DOS operating system, version 3.10
(The Microsoft software was run on the personal computers to develop code for the Intel 8088 processor on each Ziatech 8830 board.)

COMPANY ADDRESSES

Data Translation
100 Locke Drive
Marboro, MA 01752
(617) 481-3700

Microsoft Corporation
10700 Northrup Way
Bellevue, WA 98004
(206) 882-8089

Ziatech Corporation
3433 Roberto Court
San Luis Obispo, CA 93401
(805) 541-0488

SOFTWARE DEVELOPED

Prior to developing a complete system of routines for analog-to-digital conversion and data transfer out from one board, and data transfer in and digital-to-analog conversion on the other, individual functions were checked out. In all cases, software consisted of a main program, in C, calling functions (subprograms) written in 8088 assembler. All software developed is available from either Don Stilwell, mail stop SE3, NASA Johnson Space Center, Houston, Texas 77058, or from Peter Bartram, Division of Engineering, Norwich University, Northfield, Vermont 05663, by sending an MS-DOS formatted diskette.

In testing the Intel 8256 parallel interface chip (part of each board), the external interrupt was successfully used. Routines for the use of the 8256 for parallel transfer with handshake were coded first without use of interrupts by polling the handshake signals. However, the handshake pulse on the receiver side is short, independent of program instructions, and can be missed between execution of polling instructions. Therefore the interrupt signal was polled instead, even though interrupts were not enabled. Routines were also coded to make full use of the 8256 chip interrupts for handshake parallel transfer, and these were used in later work.

The Data Translation DTX328 digital-to-analog (d/a) converter used does not implement any of the interrupt capability available to modules attached to the selected peripheral processor board. Thus, for the d/a, only polling techniques were used. The DTX311 analog-to-digital (a/d) converter does allow for interrupt use to signal end-of-conversion. Separate a/d routines were prepared, one using polling only, the other making full use of the interrupt. When combined with parallel transfer, the a/d interrupt is given higher priority by the interrupt controller in the 8256 chip.

For testing system throughput, three software combinations were prepared. All three used the parallel transfer with handshake implemented with interrupts. All three used two data buffers - while one was being sent (received), the other was being prepared (used), switching back and forth between them.

The first package made no use of the analog-to-digital converter. Instead, the sending board created and sent data which represented 1024 values per cycle of a ramp

function. The receiving side used both the interrupt-based parallel transfer input and digital-to-analog conversion to output the ramp. The parallel transfer portion allowed for "extra" data to be transmitted and received - the first two bytes in each buffer were used for analog output, but additional bytes could be sent and received, though not used. This provided a means to determine how the d/a output rate would be affected by the board to board transfer.

The second software package added analog-to-digital converter support, with polling (matching the d/a on the other board). Interrupts were used for parallel transfer only. To examine the effect of conversion speeds and parallel transfer rates on system throughput, the number of a/d converted values contained in each buffer could be varied, and the number of bytes in the buffer actually transmitted could also be changed. As in the first package, only the first value in the buffer received was converted by the d/a on the other board. For example, ten samples a/d could be taken, two of them (four bytes) sent. After receiving all four bytes, only two bytes (one value) would be used for d/a output. By varying these two parameters, an examination of how the different functions affected total throughput was possible.

The third package differed from the second only in that the analog-to-digital conversion software made full use of the priority interrupt system. While this introduced higher overhead in servicing the a/d, it prevented service of the parallel transfer from locking out a/d service (except for short periods at the beginning of servicing the parallel transfer interrupt).

RESULTS

In evaluating throughput, the d/a output was monitored using an oscilloscope. For runs using both a/d and d/a converters, the analog input (from a function generator) and the parallel transfer handshake signals were also oscilloscope monitored. Conversion rates were computed from waveform periods (handshake signals) and time for "steps" in the d/a "stairstep" output. Figures presented here are accurate only in the first, occasionally second, significant digit, as the reading of the oscilloscope traces was often difficult.

For the digital-to-analog converter combined with parallel input, the maximum rate was 3600 samples output per second, which compares favorably with vendor claims. However, the conversion rate declined significantly with the receiving of additional data via parallel transfer. When four bytes were received (instead of two) and a very brief check for missing data was made using the other two bytes, output dropped to 2500 samples output per second. The rate dropped more with additional data transferred.

For the complete system, with each a/d value transmitted (two-byte buffers) and re-converted with the d/a on the other board, the throughput was 2500 samples per second without use of a/d interrupts, and between 2800 and 2900 samples per second with full use of available interrupts. However, it was shown that the analog-to-digital converter was capable of much faster rates when little parallel output was required. Rates up to about 7500 samples per second with no a/d interrupts, greater than 10000 samples per second with full interrupt use, were obtained. These rates were attained by transmitting only one sample in 100, and calculating the a/d rate from the d/a output. In this case, negligible processor time was spent in parallel transmission.

When more a/d values were transmitted to the d/a processor than were used by that board (though received), a similar decline in throughput was noted. For example, sending three a/d conversions for every one actually d/a converted resulted in an a/d conversion rate of 3700 samples per second without a/d interrupts, 4100 samples per second with a/d interrupts.

CONCLUSIONS AND RECOMMENDATIONS

As component speeds were significantly higher in isolation than when combined into a single system, it is concluded that the processor instruction time necessary to service both the real-time modules (a/d, d/a) and the parallel transfer is limiting. This is also supported by estimating the number of clock cycles needed for the necessary code compared to the number of clock cycles available between samples at desired conversion rates. Some improvement can be achieved by emphasizing timing (rather than good programming techniques) in preparing code. Operation of the parallel transfer without interrupts would save the overhead time of processing the interrupt (register saves and restores, etc.), for example. While using interrupts for the a/d increased throughput, it did so in spite of more instruction clock cycles required, by preventing the parallel transfer from taking priority over the a/d service. Use of real-time hardware modules which do not require as much software to service would also give some improvement. Substitution of a faster processor, such as the NEC V20, which is pin compatible with the Intel 8088 but five to ten per cent faster for the same clock rate, would also give marginal improvement.

More significant gains require removing some of the input/output overhead. For board to board transfer, the STD bus could be used instead of the 8256 interface. This would require the transfer to be programmed through the master processor, as the peripheral processor boards cannot control the STD bus. This would remove much of the parallel transfer code from the peripheral processors, but the master would no longer be totally dedicated to the principal investigator. Because direct memory access (DMA) is available on the master board considered, the overhead for the master would be much less than on the peripheral processor boards. Alternatively, DMA could be added to the peripheral processor boards. The possibility of designing a "piggyback" module adding appropiate DMA hardware is worth investigating. By reducing the involvement of the processor in moving data between memory and the 8256 parallel interface chip for board to board transfer, the processor would have more time to service the real-time modules at design conversion rates.

The potential is great for major gains over the present

LSLE microcomputer. The multiple-processor approach is still the most attractive, because of the conceptual advantages of sharing work among many units. Further study should provide the means to improve significantly the data throughput rates.

HEURISTIC CONTROL

OF THE UTAH/M.I.T. DEXTROUS ROBOT HAND

Final Report

NASA/ASEE Summer Faculty Fellowship Program -- 1987

Johnson Space Center

Prepared by:	**Andrew H. Bass, Jr.**
Academic Rank:	**Assistant Professor**
University & Department:	**Texas Southern University Department of Computing and Applied Sciences Houston, Texas 77004**
NASA/JSC	
Directorate:	**Engineering**
Division:	**Systems Development and Simulation Division**
Branch:	**Intelligent Systems Branch**
JSC Colleague:	**Kathleen J. Healey**
Date:	**August 26, 1987**
Contract Number:	**NGT 44-001-800**

ABSTRACT

The main thrust of this work has been to analyze basic hand grips and sensor interactions that a dextrous robot hand will need as part of the operation of an EVA-Retriever.

The work focused on understanding what is to be done with a dextrous robot hand and how such a complex machine might be controlled. It was assumed throughout that an anthropomorphic robot hand should perform tasks just as a human would; that is to say, at least initially, the most efficient approach to developing control strategies for the hand would be to model actual human hand actions and do the same tasks in the same ways[1]. Therefore, a Heuristic approach to control was developed.

The tasks performed by human hands are extremely complex involving the movement of many fingers and joints and contact between numerous surfaces, that is, contact at an infinity of points. In addition, multiple sensors including force/touch, vibration, shear and sometimes heat must be taken into account. Each movement and task involves a multitude of these force/touch interactions and a number of sub-sequences and events, the aggregate of which, can only be termed a skill. In addition, I have tried to understand just what is to be done with a dextrous hand. Therefore, basic hand grips that human hands perform, as well as hand grip actions were analyzed.

It was also important to examine what is termed *sensor fusion*. This is the integration of various disparate sensor feedback paths. These feedback paths can be spatially and temporally separated, as well as, of different sensor types. Neural Networks[2] are seen as a means of integrating these varied sensor inputs and types. Basic Heuristics of hand actions and grips were developed. These Heuristics offer promise of control of dextrous robot hands in a more natural and efficient way. Emphasis was also placed on possible methods of implementing these techniques. Future work will be to continue development of routines and Heuristics and use them to control the Utah/M.I.T. dextrous robot hand. A smart robot hand, one that can adapt to new situations and develop new skills is a goal of researchers working with dextrous robot hands. It is an ultimate goal that a smart hand will be able to develop its own Heuristics which will make its operation and use even more efficient and its control more simple.

Introduction

The EVA-Retriever

The EVA-R will evolve into a completely autonomous mobile robot. The robot will be used for retrieving objects which have floated free from Shuttle or Station and will secure or stabilize objects as need be. EVA-R can be used to assemble structures and as a general Astronaut's assistant. The robot will have a vision system and possibly some form of laser ranging device. It is expected that the EVA-R will be able to recognize tools and other objects and upon voice command locate, identify, track and retrieve them. EVA-R will ultimately use tools to perform work assignments.

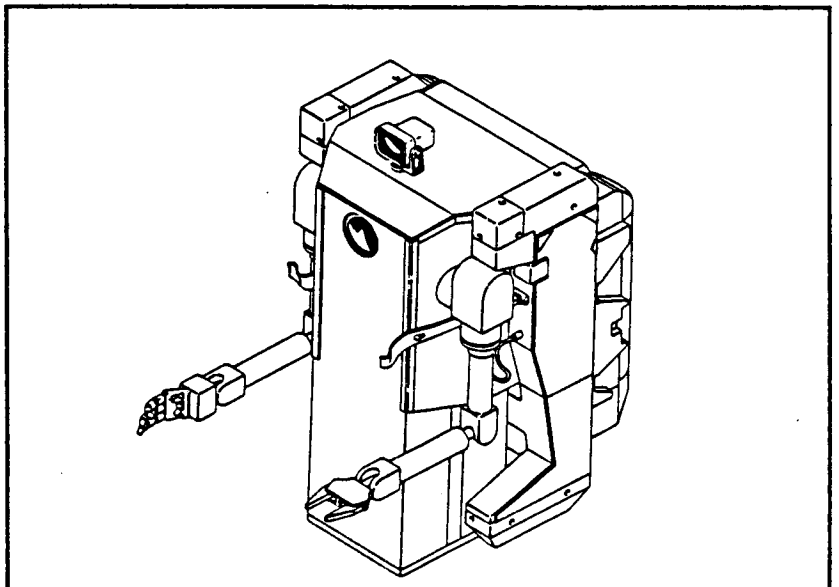


Figure 1. Extra Vehicular Activity-Retriever (EVA-R)

As can be seen in figure 1., the EVA-R will reside in an MMU and will have two arms. At least One arm will have a smart dextrous hand like the Utah/M.I.T. hand.

An autonomous mobile robot will require several new and emerging technologies. These include :

1. Neural Networks which have been trained to identify and interpret sensory input.
2. AI languages such as CLIPS that can interface with the neural networks and be programmed to deliver higher level commands to the robot and aid in task planning and execution.
3. Dynamic adaptive task planners[3,4] which can make decisions based on sensory information, task safety, present and evolving conditions, changing goals, and previous experience.

4. The robot must have the ability to learn from its experiences and discard useless information. This type of learning by the robot may possibly be implemented through neural networks.
5. Dextrous hands for using tools and performing skilled tasks.
6. Touch sensors and vibration/feeling sensors incorporated into the hand at appropriate places.
7. Vision sensors and ranging sensors for navigation and identification of objects.
8. Voice Recognition and understanding so that it can carry out command given by an Astronaut.

Development of Neural Network technology is key to the the implementation of a fully functioning autonomous mobile robot. Whether the neural networks are hard-wired or simulated they will require faster chip computational speeds and large amounts of nonvolatile memory. The various components of the robot must have separate and parallel networks. The hand will have neural networks for sensing which will be distinct from the vision networks and voice recognition networks. A guidance navigation and balance network would be used to integrate and interpret the various outputs of the other nets. Layered between the networks would be an AI language such as CLIPS. A top or overseeing AI layer would manage the robots activities.

The Utah M.I.T. Dextrous Robot Hand

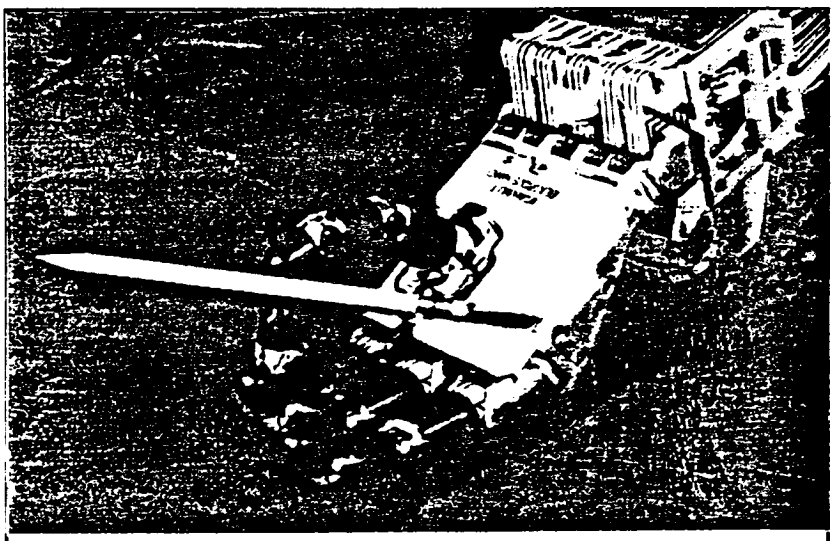


Figure 2. The Utah/M.I.T. Dextrous Robot Hand

The Utah/M.I.T. dextrous robot hand is a four fingered hand with each finger having multiple joints. This includes one thumb which opposes the three fingers. There are three joints in each finger which serve to curl the finger, and one base joint which rotates the entire finger back and forth in the plane of the palm. The hand is capable of graceful and delicate movement just as a human hand is and, therefore, can perform many human hand actions.

Findings

Kinematics

To study control strategies and simulate hand actions a kinematic solution of the hand was sought. A typical finger is considered and can be referred to a single point or coordinate system in the palm. Using the Denavit-Hartenberg notation, the position and orientation fingertip is expressed as a series of matrix transformations.

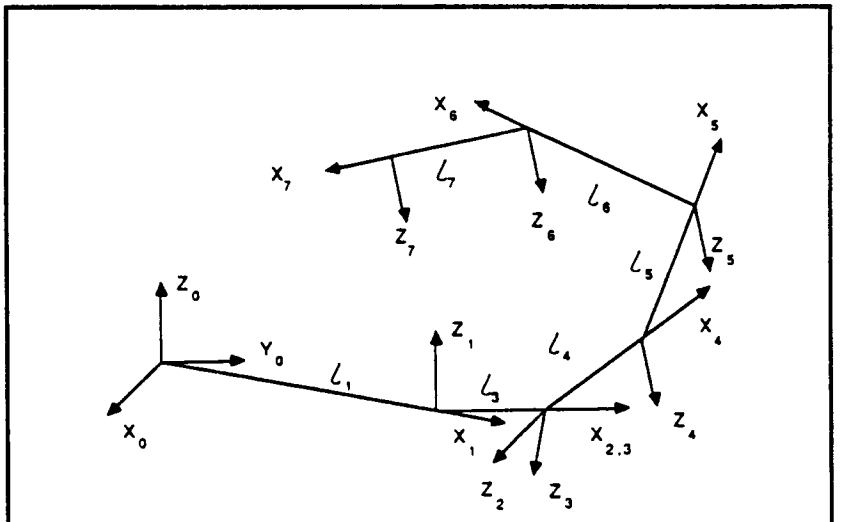


Figure 3. FINGER KINEMATIC STRUCTURE

The transformation from the reference coordinate system (X_0,Y_0,Z_0) to a typical fingertip (X_7,Y_7,Z_7) is given by

The fingertips position and orientation are given by :

(1.) $$T = A_1^0 A_2^1 A_3^2 A_4^3 A_5^4 A_6^5 A_7^6$$

A_i^{i-1} are matrix transformations that give the coordinate system at point i along link l_i from i-1th joint i-1 to ith joint [5,6]. Since A_1^0 and A_2^1 are constant matrices let

(2.) $$T^* = (A_1^0 A_2^1)^{-1} T$$

(3.) $$T^* = \begin{pmatrix} n_x & t_x & b_x & P_x \\ n_y & t_y & b_y & P_y \\ n_z & t_z & b_z & P_z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The upper left hand 3x3 contains the direction cosines that give the orientation of the fingertip and the first three elements of the fourth column give the position of the fingertip from the (X_2,Y_2,Z_2) coordinate system.

Using the Denavit-Hartenberg Notation a homogeneous transformation from the i-1th joint to the ith joint along link l_i is :

$$A_i^{i-1} = \begin{pmatrix} \cos\theta_i & -\sin\theta_i\cos\alpha_i & \sin\theta_i\sin\alpha_i & a_i\cos\theta_i \\ \sin\theta_i & \cos\theta_i\cos\alpha_i & -\cos\theta_i\sin\alpha_i & a_i\sin\theta_i \\ 0 & \sin\alpha_i & \cos\alpha_i & d_i \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

where θ_i is a rotation of the revolute joint i, d_i is for a prismatic joint (here = 0), a_i is the link length, and α_i is the link twist. For a typical finger

$$A_1^0 = \begin{pmatrix} \cos\theta_1 & -\sin\theta_1 & 0 & l_1\cos\theta_1 \\ \sin\theta_1 & \cos\theta_1 & 0 & l_1\sin\theta_1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

θ_1,θ_2 are constant. $\alpha_1 = 0, \alpha_2 = \pi/2$ and $l_2 = 0$. For convenience $\theta_3 = 0$. Then $\alpha_3 = \theta_3$. Also, $\theta_4 \approx 30°$ and $\alpha_4 = 0$. Therefore, A_4^3 is a constant matrix.

$$A_2^1 = \begin{pmatrix} \cos\theta_2 & 0 & \sin\theta_2 & 0 \\ \sin\theta_2 & 0 & -\cos\theta_2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad A_3^2 = \begin{pmatrix} 1 & 0 & 0 & l_3 \\ 0 & \cos\theta_3 & -\sin\theta_3 & 0 \\ 0 & \sin\theta_3 & \cos\theta_3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and for i = 4,5,6,7 the transformations and their inverses are [7]

$$A_i^{i-1} = \begin{pmatrix} \cos\theta_i & -\sin\theta_i & 0 & l_i\cos\theta_i \\ \sin\theta_i & \cos\theta_i & 0 & l_i\sin\theta_i \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad (A_i^{i-1})^{-1} = \begin{pmatrix} \cos\theta_i & \sin\theta_i & 0 & -l_i \\ -\sin\theta_i & \cos\theta_i & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

from (2.)

$$T^* = A_3^2 A_4^3 A_5^4 A_6^5 A_7^6$$

Postmultiplying both sides by $(A_7^6)^{-1}$ yields.

$$T^*(A_7^6)^{-1} = A_3^2 A_4^3 A_5^4 A_6^5$$

Premultiplying both sides of the equation by $(A_3^2)^{-1}$ equating both sides and solving for θ_3.

$$(A_3^2)^{-1} T^* (A_7^6)^{-1} = A_4^3 A_5^4 A_6^5$$

$$\cos\theta_3(P_x - l_7 n_x) - \sin\theta_3(P_y - l_7 n_y) = 0$$

$$\theta_3 = \tan^{-1}\left(\frac{P_x - l_7 n_x}{P_y - l_7 n_y}\right)$$

Solving for θ_5 . Let

$$W = (A_3^2 A_4^3)^{-1} T^* = A_5^4 A_6^5 A_7^6$$

$$(A_5^4)^{-1} W = A_6^5 A_7^6$$

Equating sides evaluating and solving yields :

$$\theta_5 = \tan^{-1}\left(\frac{W_{23}}{W_{13}}\right)$$

Solving for θ_6,θ_7 in a similar fashion.

$$(A_6^5)^{-1} (A_5^4)^{-1} W = A_7^6$$

$$\theta_6 = \tan^{-1}\left(\frac{W_{13}\cos\theta_5 + W_{23}\sin\theta_5}{W_{13}\sin\theta_5 + W_{23}\cos\theta_5}\right)$$

Taking the ratio of the 2,4 term to the 1,4 term and solving yields :

$$\theta_7 = \tan^{-1}\left(\frac{P_y(\cos\theta_5\cos\theta_6 - \sin\theta_5\sin\theta_6) - P_x(\cos\theta_5\sin\theta_6 + \sin\theta_5\cos\theta_6) + l_5\sin\theta_6}{P_x(\cos\theta_5\cos\theta_6 - \sin\theta_5\sin\theta_6) + P_y(\cos\theta_5\sin\theta_6 + \sin\theta_5\cos\theta_6) - l_5\cos\theta_6 - l_6}\right)$$

In the above relations :

$$W_{13} = b_x \cos\theta_4 + b_y \cos\theta_3 \sin\theta_4 + b_z \sin\theta_3 \sin\theta_4$$

$$W_{23} = -b_x \sin\theta_4 + b_y \cos\theta_3 \cos\theta_4 + b_z \sin\theta_3 \cos\theta_4$$

Transforming back to the base coordinate system (X_0,Y_0,Z_0) from the (X_2,Y_2,Z_2) coordinate system.

$$B_x = \cos\theta_1(B_{x0}\cos\theta_2 + B_{y0}\sin\theta_2) + \sin\theta_1(B_{y0}\cos\theta_2 - B_{x0}\sin\theta_2)$$

$$B_z = \cos\theta_1(B_{x0}\sin\theta_2 - B_{y0}\cos\theta_2) + \sin\theta_1(B_{x0}\cos\theta_2 + B_{y0}\sin\theta_2)$$

$$P_x = \cos\theta_1(P_{x0}\cos\theta_2 + P_{y0}\sin\theta_2) + \sin\theta_1(P_{y0}\cos\theta_2 - P_{x0}\sin\theta_2) - l_1\cos\theta_2$$

$$P_z = \cos\theta_1(P_{x0}\sin\theta_2 - P_{y0}\cos\theta_2) + \sin\theta_1(P_{x0}\cos\theta_2 + P_{y0}\sin\theta_2) - l_1\sin\theta_2$$

$$N_x = \cos\theta_1(N_{x0}\cos\theta_2 + N_{y0}\sin\theta_2) + \sin\theta_1(N_{y0}\cos\theta_2 - N_{x0}\sin\theta_2)$$

$$N_z = \cos\theta_1(N_{x0}\sin\theta_2 - N_{y0}\cos\theta_2) + \sin\theta_1(N_{x0}\cos\theta_2 + N_{y0}\sin\theta_2)$$

$$B_y = B_{z0} \quad , \quad N_y = N_{z0} \quad , \quad P_y = P_{z0}$$

The Thumb is opposite the fingers operates in an opposing direction.
In this case :

$\theta_2 = \xi_{32} + \pi, \quad \theta_4 = \pi/2.$

Traditional Kinematics Versus the Heuristic Approach

Traditional mathematical approaches involve dynamic modeling with the equations of motion. These models solve for end effector orientation, position, motions, and forces; also joint angles and torques. This is feasible for robots and machines with few degrees of freedom and simple geometries. However, Even for relatively simple robot mechanisms, the mathematics can be intractable. The complexity of a dextrous robot hand and the need to include smart and adaptive control suggests the use of heuristic methods. Asada and

Slotine[5] page 185 state, " ... very few manipulators are capable of such seemingly simple tasks as driving a screw or turning a crank." They were discussing a class of problem called Compliant Motion control which they define as " ... concerned with the control of a robot in contact with its "environment" - an object to manipulate or assemble ...".

The difficulties of compliant motion control when there are only a few points of contact between the manipulator and the work piece further point out the futility of using traditional mathematical approaches. This is especially true when a dextrous hand is concerned and there are an infinite number of points, indeed surfaces, which are in contact with an object being manipulated. A purely mathematical approach would require major breakthroughs and brand new insights.

As further evidence, Jacobsen, co-inventor of the Utah/MIT hand [13], in his speech to the Space Operations Automation and Robotics SOAR Conference at NASA/JSC August 6, 1987 stated of the dextrous hand he had created, "we really don't know how to control these things". He went on to state that you must understand what it is you really want to do with them. For example, what is a grip and how do you do various grasps.

Insight into the efficient performance of manipulator tasks is easily obtained by close observation of ones own manual skills and abilities. The proper placement and number of contact sensors is important. It is also of primary importance to examine the issues of what is termed *sensor fusion*. This is the integration of various disparate sensor feedback paths, spatially, temporally and sensor type (touch and vision for example).

The Heuristic Approach

Webster defines heuristic as follows : Serving to guide, discover, reveal. Valuable for stimulating or conducting empirical research but unproved or incapable of proof.

Used here, it is the development of AI techniques to control ROBOT devices. Heuristic approaches offer advantages in that they allow controlling the robot hand in human-like, complex operations and motions which are difficult if not impossible to describe in a mathematical fashion at the present time. They involve developing rules of thumb and implementing them in an AI sense. This does not rule out the melding of traditional control and mathematical structure with AI-control,learning and expert system supervision[9]. There could be control on a machine or low level by kinematic and dynamic analysis with Robustness and flexibility supplied by expert systems. An implementation might involve the use of neural networks for low (machine) level control. Neural nets would develop skills in performing tasks in a fashion similar to the development of dextrous skills in a human. The ultimate test of a particular implementation will be the ability of the robot to perform tasks in new and unfamiliar situations and the ability to make judgement type decisions when needed.

Classification of Hand Actions

An understanding of how a hand performs a particular task is gained by understanding what those tasks are. Hand Actions are developed as a means of classifying the various tasks to be performed by the dextrous robot hand.
After an object has been gripped, there are many actions that are possible. A few basic actions that the hand will have to perform are listed below.

- Holding : Maintenance of a grip
- Pulling : Grip + Holding + Force in the direction of the wrist
- Pushing : Contact with an object + a Force away from the wrist
- Feeling : Contact with an object to discover its characteristics.
- Rotating : Grip + Wrist rotation accompanied by a possible sinusoidal movement of the fingers.
- Turning (wrench) : Grip + Force in the direction of the palm and perpendicular to the wrist axis.

Common among hand actions are grips. To understand how to perform the various hand actions listed above it is necessary to first study grips and grip primitives. These are basic contact configurations that a dextrous hand must be capable of performing.

Classifications of GRIPS

- Gripping : is the act of securing an object with the hand.
- Grip Primitive : A grip which is used to hold an elementary object or object type and is distinguishable from other grips by a unique geometric pattern of the fingers.
- Elementary Object : From a gripping point of view , all objects are composed of elementary objects such as cylinders, spheres,large planar surfaces, etc.

Grips can be divided into states that are generally distinguishable from each other by their geometry, function or some combination thereof. A grip for an object type can be scaled within limits as a function of object size. A heuristic can handle the break down of scaling.

Spherical and Cylindrical Grips

ELEMENTARY GRIP PATTERN CHART					
OBJECT	FINGER/JOINT in °	0	1	2	3
CYLINDER	THUMB	0	ϕ	ϕ	ϕ
ROD	FINGER 1	ϕ	"	"	"
	FINGER 2	"	"	"	"
	FINGER 3	"	"	"	"
CYLINDER	THUMB	45*	0	0	0
LEVER	FINGER 1	ϕ	ϕ	ϕ	ϕ
	FINGER 2	"	"	"	"
	FINGER 3	"	"	"	"
SPHERE	THUMB	0	ϕ	ϕ	ϕ
	FINGER 1	$-\mu$	"	"	"
	FINGER 2	0	"	"	"
	FINGER 3	μ	"	"	"

$\mu = \max \angle$, $\phi = 2\tan^{-1}(l/2r)$

The angles ϕ are calculated with equal length finger links. Actual angles can be determined using the kinematics calculations or by measurement with the robot hand. The chart above serves to illustrate the general relationship between the angles. *In a Spherical or Cylindrical Grip it is the thumb and the 0 joint that give a grip its unique configuration.*

*For a cylindrical lever grip, the thumb is rotated along the axis of the shaft. Presently, the Utah/M.I.T. hand cannot rotate the full 90 degrees to accomplish this. The hand can only rotate $\approx \pm 45°$ from the normal to the plane of the palm. A future version of the hand will be capable of rotating $\pm 90°$: (personal conversation with Jacobsen).

Flat Surface Grips

A flat object, from a one handed gripping perspective, is one in which the only grip opportunity arises on a side of the object and the fingers and thumb are in contact with opposite parallel faces of the object. There are generally three sub categories of grippable flat objects.

FLAT SURFACE GRIP PATTERN CHART					
OBJECT	FINGER/JOINT in °	0	1	2	3
SMALL EDGE	THUMB	0	0	0	0
non-massive	FINGER 1	0	ϕ	ϕ	ϕ
(FINGERTIP)	FINGER 2	0	"	"	"
	FINGER 3	0	"	"	"
SMALL EDGE	THUMB	90	0	0	0
non-massive	FINGER 1	0	ϕ	ϕ	ϕ
(PINCH)	FINGER 2	0	"	"	"
	FINGER 3	0	"	"	"
massive	THUMB	90	0	0	0
(PALM GRIP)	FINGER 1	0	90	0	0
depends on	FINGER 2	0	90	0	0
edge width	FINGER 3	0	90	0	0
LARGE EDGE	THUMB	0	0	0	90
(Largest	FINGER 1	0	0	0	90
Grasp)	FINGER 2	0	0	0	90
	FINGER 3	0	0	0	90

1. Small Edge Non-Massive object: The edge width is less than the distance from thumb base to finger base. In this case a fingertip grip is used. Another grip can be used in this case, with the thumb and first finger. The sheet rests against the outer side of the first finger and presses or pinches the sheet against this finger. The remaining fingers are closed against the palm to give support to the first finger.
2. Small Edge Massive object: The edge width is less than the distance from thumb base to finger base. The object is massive relative to hand strength so that a firm grip is needed. The palm of the hand is used as a support platform and the fingers secure the object. To do this the thumb is rotated 90° so that it aids in the platform that the palm makes. This platform is placed on one side of the flat object. The fingers then wrap around the edge of the object. One corner of the edge (the corner farthest away from the palm) rests in a 90° bend formed by the finger joints. This bend is formed as close to the palm as possible (for strength of grip) at the same joint level in all fingers.
3. Large Edge: The edge width is greater than the distance from thumb base to finger base but less than the distance between the outer most finger joints and thumb joints when all joints are at 0 angle. In this case, a minimum distance between the object edge and the palm of the hand and maximum finger and thumb surface area contact with the object is sought.

Altered Grips

These are grips that are used with any object ,even an elementary object for which there is a grip primitive. They are used to grip the object in a different fashion to serve a different purpose. For example, a screw driver is basically a cylinder but when this tool is to be used, a cylindrical grip primitive is altered so that a force along the axis of the screw driver can be applied and a wrist turning motion along the arms axis is performed. An altered grip can be thought of as a task grip.

Task or Tool Grips

A tool from a gripping perspective is composed of elementary objects (see Classification of Hand Actions). However, when a tool is used the grips that are used are not elementary grips but are grips that take into account the forces and motions that are needed to perform a task.

Most EVA tools have cylindrical handles or handles that can be gripped with a cylindrical grip [10,11].

The tool generally lays across the palm of the hand. For example a flashlight, hammer or screwdriver. An automatic screwdriver or drill which has a gun configuration is used with a cylindrical type grip. The trigger is actuated by the first finger which uses a variable cylinder configuration.

TASK/TOOL GRIP PATTERN CHART					
OBJECT	**FINGER/JOINT in °**	**0**	**1**	**2**	**3**
SCREW DRIVER	THUMB	90	0	0	0
FLASH LIGHT	FINGER 1	0	.5 ϕ	ϕ	ϕ
HAMMER	FINGER 2	ϕ	.75 ϕ	"	"
	FINGER 3	ϕ	ϕ	"	"
ELECTRIC	THUMB	0	0	0	0
DRILL	FINGER 1		0	X	X
	FINGER 2	0	ϕ	ϕ	ϕ
	FINGER 3	0	"	"	"

In the Electric Drill grip X is a variable angle representing the pulling of the trigger. $0 \leq X \leq \phi$.

Interpolated Grips

This type of grip is a combination of other grips (Elementary/Task) and is used to create or compose a new grip of an object whose grip is not known. An example of an interpolated grip would be a manual gear shift lever in an automobile that is a long cylindrical shaft with a spherical knob on top.

INTERPOLATED GRIP PATTERN CHART					
OBJECT	**FINGER/JOINT in °**	**0**	**1**	**2**	**3**
GEAR SHIFT	THUMB	90	0	0	0
(Spherical	FINGER 1	-μ	ϕ	ϕ	ϕ
top)	FINGER 2	0	ϕ	"	"
	FINGER 3	0	ϕ	"	"
VALVE-	THUMB	-μ	ϕ	ϕ	ϕ
HANDLE or	FINGER 1	-μ	ϕ	ϕ	ϕ
DOOR KNOB	FINGER 2	0	2ϕ	2ϕ	2ϕ
(Spherical)	FINGER 3	90	90	90	90

These are a two examples of interpolated grips. It is necessary to develop an Interpolated Grip Strategy in order to create a truly smart hand.

Location of Sensors

LOCATION of CONTACT SENSORS CHART					
GRIP	FINGER/LINK	PALM	1	2	3
FLAT SURFACE	THUMB	0	1	1	1
SMALL EDGE	FINGER 1	0	0	0	1
(FINGERTIP)	FINGER 2	0	0	0	1
	FINGER 3	0	0	0	1
(PINCH)	THUMB	0	1	1	1
	FINGER 1	0	side	side	side
	FINGER 2	0	0	0	0
	FINGER 3	0	0	0	0
SPHERE or	THUMB	1	1	1	1
CYLINDER	FINGER 1	1	1	1	1
	FINGER 2	1	1	1	1
	FINGER 3	1	1	1	1
VALVE	THUMB	1	1	1	1
DOOR KNOB	FINGER 1	1	1	1	1
	FINGER 2	1	side	side	side
	FINGER 3	0	0	0	0

A necessary condition to perform various human like grips and tasks is the use of contact sensors. Particular grips require the placement of touch sensors in specific locations. The chart above was created using grips from several other charts. A zero indicates no sensor needed and one indicates the need for a sensor at that location. Although only one chart is given the chart indicates that a contact sensor is needed at each link and also in the palm at the base of each finger. In addition, sensors are needed along two of the fingers. Finger one requires sensors along the outside (right side looking toward the palm) at every link and finger two will require sensors on the same side as finger one on the two end links.

Grip Alteration Action

A *Grip Alteration Action* is used to change from one grip to another. The reasons for doing this are to position the object differently in the hand for a particular purpose. Those purposes could range from attaining a more secure grip to a specific grip that is used with a tool. Some tools require a changing grip for their use such as an automatic drill/screw driver. Using a hammer on earth requires grip alteration during usage. A hammer grip slides on the hammer shaft allowing gravitational effect plus the lever effect to increase its kinetic energy. In the case of a tool, there may be several grip alteration actions that are used.

Clips and C

CLIPS is a forward chaining language developed at NASA/JSC[12,13,14] by the AI section of the Mission Planning and Analysis division. It is written in the C language and is closely associated with C. C programs can be called from CLIPS and C programs can call CLIPS routines. It also has interfaces to ADA and Fortran although these are not as easy as with C. The language is easy to learn and affords easy implementation of pattern matching.

Heuristic Control Using CLIPS

As an example of a control structure using CLIPS

```
(defrule grab-handrail
    ?f1 <- (close-hand standard-handle1 ?speed)
        ;If the pattern above matches a fact that had been
        ;previously asserted, this rule will be put on the
        ;agenda and will be ready to fire. Also, the variable
        ;?speed is instantiated with the value located at that
        ;position in the pattern.
=>
    (retract ?f1)
    (assert (status good-grab = (grab standard-handle1 ?speed))))
        ;grab is a c routine that does the actual hand closing.
        ;status good-grab is asserted as a new fact.
```

Neural Networks and C

A Neural Network Simulator written in C is being developed at NASA/JSC[15] by the AI section of the Mission Planning and Analysis Division. C routines and therefore, CLIPS routines should be callable from this environment and visa versa. At present the networks are back propagation.

A spatial-temporal neural network* could be used to fuse and integrate sensory feedback in the hand. Given point contact sensors, as opposed to an array of sensors, one at each link and one or more in the palm of the hand a simple spatial-temporal network could be developed and trained to recognize space-time patterns of contact. If the space-time pattern of contact *approximates* a pattern that had been previously learned then that pattern would be recognized and further action could or could not be taken based on that outcome. For example, there are typical space-time contact patterns that arise in the

gripping of a cylinder. Assuming that the object had been previously identified as a cylinder a particular space-time contact pattern would confirm that the proper grip on the cylinder had been made. That pattern, might first include palm contact, followed by base finger link contacts and so on, out to the fingertips. If a neural network implementation operated in real time a deviation from that pattern during grip execution could possibly signal that corrective action is necessary to complete and secure the grip. In addition, given a real time network, palm contact could signal to the controlling software to begin the cylindrical grip. Alternatively, the actual grip could be implemented by a neural network that had been trained to respond to patterns of contact input.

* J. Freeman, Ford Aerospace, Johnson Space Center, personal conversation.

Heuristic Grip Summary

1. A Cylindrical grip primitive can be used for most objects.
2. Grip as close to the center of mass as possible[16].
3. If the object's characteristic length (in the grip region) is $L_{object} < 3/4 L_{hand}$ the distance from the wrist to the fingertips, a one-handed grip will probably work.
4. For objects with characteristic lengths (in the grip region) $L_{object} < l_{tip} + l_{tip-1}$ the two outer finger link lengths and $F_{hand}/M_{object} \ll 1$ a fingertip grip can be used.
3. Use hand sensors touch and vibration to determine if grip is slipping.
4. Touch sensors are used to signal the beginning of a grip action.
5. Use vision sensors for sufficient condition confirmation that a grip has been successfully made.
6. Plan Grip Action Script and continuously monitor safety during execution.
7. In general it is the thumb and 0 joint that give a grip its unique configuration and therefore function.

Conclusions

Future Work and Recommendations

The work reported in this paper is by no means exhaustive. It serves to illuminate a path that can be followed with what I feel will be real tangible results and successes in the development of an autonomous robot with a dextrous hand. The fabric which ties all of the components together will be the development and application of Neural Networks to sensors and learning, further development of an AI language such as CLIPS that can easily interface the Neural Networks, and the development of a dynamic adaptive task planning strategy. Also, it is necessary to develop an Interpolated Grip Strategy in order to create a truly smart hand. This would be part of the dynamic adaptive task planner.

Recommendations

Sensors : Contact sensor are needed at each link and also in the palm at the base of each finger. Sensors are needed along the side of two of the fingers.

Neural Network research should be applied the sensor fusion problem.

REFERENCES

1. S.C. Jacobsen,E.K. Iverson,D.F. Knutti,R.T. Johnson,K.B. Biggers, "Design of The Utah/M.I.T. Dextrous Hand" Submitted for presentation at the 1986 *IEEE* International Conference on Robotics and Automation, April 7-10, 1986 San Francisco, California
2. R.P. Lippman, " An Introduction to Computing With Neural Nets, *IEEE* ASSP Magazine, April 1987
3. Y. Ichikawa,N. Ozaki, "A Heuristic Planner and an Executive for Mobile Robot Control", *IEEE* Transactions on Systems, Man. and Cybernetics, Vol. SMC-15, No. 4, July/August 1985
4. D. Chapman, "Planning for Conjunctive Goals", *Artificial Intelligence*,32 (1987), 333-377
5. H. Asada,J,-J.E. Slotine, "Robot Analysis and Control", John Wiley & Sons, Copyright 1986
6. A. Hemami, "On a Human-Arm-Like Mechanical Manipulator", Department of Mechanical Engineering, Concordia University, SGW Campus-Annex B308, 1455 de Maisoneuve Blvd. West, Montreal, Quebec H3G 1M8 (Canada), January 1986
7. R.P. Paul, B. Shimano, G.E. Mayer, "Kinematic Control Equations for Simple Manipulators", *IEEE* Trnasactions on Systems, Man, and Cybernetics, Vol. SMC-11,No. 6, June 1981
8. K.B. Biggers,S.C. Jacobsen,G.E. Gerpheide, "Low Level Control of the Utah /M.I.T. Dextrous Hand" *IEEE* 1986
9. C.T. Kitzmiller,J.S. Kowalik, "Coupling Symbolic and Numeric Computing in Knowledge-Based Systems", AI Magazine, Summer 1987
10. "EVA Catalog Tools and Equipment" NASA/JSC November 4,1985 JSC-20466
11. C.E. Whitsett, "New Tools for EVA Operations", 17Th. Intersociety Conference on Environmental Systems Seattle, Washington July 13-15, 1987 SAE Technical Paper Series 8791499
12. "CLIPS User's Guide" JSC-22308 October 1986
13. "CLIPS Reference Manual" JSC-22552 Version 4.0 March 1987
14. Dr. Joseph Giaratano , Chris Culbert, Gary Riley, Robert T. Savely, "A Solution to the Expert System Delivery Problem" NASA/JSC Draft 6/8/87
15. Robert T. Savely, Editor "The Implementation of Neural Network Technology" Presented at the First Annual International Conference on Neural Networks San Diego, California June 21-24 1987
16. J.D. Wolter, R.A. Volz, A.C. Woo, "Automatic Generation of Gripping Positions" *IEEE* Transactions on Systems, Man, and Cybernetics, Vol. SMC-15, No. 2, March/April 1985

A REVIEW OF HYPERTEXT IN A NASA PROJECT MANAGEMENT CONTEXT

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1987

Johnson Space Center

Prepared by:	Christopher J. Bell
Academic Rank:	Assistant Professor
University & Department:	Illinois State University Applied Computer Science Dept. Normal, Illinois 61761
NASA/JSC	
Directorate:	Space and Life Sciences
Division:	Man-Systems
Branch:	Crew Station
JSC Colleague:	Marianne Rudisill
Date:	August 15, 1987
Contract Number:	NGT 44-001-800

Abstract

This document discusses the principles of data storage, the comparative strengths of data bases, and the evolution of hypertext within this context. A classification schema of indexing and of hypertext document structures is provided. Issues associated with hypertext implementation are discussed and potential areas for further research are indicated.

Introduction

One of the strengths of the modern computer is its ability to manipulate large masses of data. It is anticipated that the space station will carry on board a great deal of the documentation associated with the design and construction of its components. Accessing needed information from data sets of the anticipated size of the one planned for the space station is likely to create difficulties for the user. The user may find it extremely difficult to locate the desired information or may find that the information sought is contained in a large mass of irrelevant material. Further, the ability to efficiently access most modern data bases is limited by the familiarity of the user with the peculiarities of the database organization.

Hypertext is a method for organizing text oriented data bases to facilitate ease of access, to promote rapid navigation to desired nodes, and to provide views of data paths that facilitate consideration of alternatives which might be of interest for browsing. Users find hypertext based documents facilitate browsing and simplify navigation problems that can become quite severe in other data search environments.

This paper will examine the data processing/file manipulating background against which the strengths of hypertext concepts should be viewed. Several hypertext implementations are reviewed. The author has worked extensively with one of these packages (HOUDINI by MaxThink) on a Compac Deskpro (IBM-PC XT equivalent), and some of the observations given late in the paper are the result of several implementations developed by the author using the above combination. Areas of research in human-computer interaction using hypertext that should be addressed are described.

Historic Development/Overview of Varieties of Databases

The computer has come to be viewed as a necessity in the modern world for many reasons - speed of processing, significant improvements in numeric manipulations, ability to manage large volumes of data, among many other strengths. The ability to manipulate large masses of data, to perform the same process to an endless number of records or to search a large file for the specific record needed has made the computer a mainstay of the information processing world. The first section of this paper will provide a review of record, file, and database fundamentals common to the information processing environment, and attempt to

provide the reader with the background useful to understanding the virtues of hypertext.

The primary structure of information is the field. A field consists of the smallest segment of data meaningful to the user. Examples of fields might include name, home address, work address, etc. Fields alone are meaningless unless they belong to a higher data organization called a record. Each record of the same type will contain the same fields, but the contents of the fields will differ from one instance of a record to another. Associated records are grouped into files which serve as the primary organizational structure for large sets of data. An organization may maintain separate files for its employees, for its corporate resources, for its customers or students, etc.

Traditionally we have tried to organize these files into one of three structures based on the anticipated accesses which will be made to the data. If access will be to most or all of the records during any given file use, and the order of the accesses is not important or can be anticipated (i.e. alphabetical), then a simple sequential structure where one record follows immediately upon the previous will make the most efficient use of storage and reading time. Sequential storage is still the classic data organization and is found useful in many situations, but it has major deficiencies for some situations.

Many users find that they cannot predict in advance which records they are going to need or how frequently they will need to access any given record. When this is the case, files can be organized so that the location of any given record can be determined from a unique key field (such as part number) contained in each record. Thus an attempt to access a given record can be directed to the location of that record in storage, with no need to process or pass over unused records. This direct access method provides much faster access to any given record, but at the cost of wasted storage space (for locations with no corresponding record key field) and no easy way to access all records if the user so desires.

If these latter considerations are important to the user's anticipated application, one may organize the file so the records are stored immediately following one another - as in sequential - but one also maintains an index pairing the record's key field with storage location, one can have most of the storage efficiency of sequential organization with rapid access to any individual record as found with direct organizations. This indexed sequential structure provides both of these boons, but with some cost associated with maintaining and accessing the index. Most of the currently available products, even those

specific to certain manufacturers (such as IBM's KSDS) can be categorized into one of these formats.

Common to all basic file organizations, however, are problems related to multiple files of similar data. As each user will have needs for data which differ somewhat from other users, each will require a file dedicated to them. Thus data is often repeated in several files, wasting space and resources. This redundancy may lead to much worse problems caused by inconsistency among files containing duplicate record fields. With many users accessing many different files, it becomes increasingly hard to maintain corporate standards of security and data integrity. Laws to control what may or may not be retained must be respected, and with diverse users and files, it becomes difficult to do so. Each attempt to initiate a new use of the computer system requires extreme investments in duplicating file services, leading to greater and greater demands on the system resources. Further if the company should wish to begin using a new computer system, file migration problems must be addressed for each application file, duplicating expenses, if even resolvable.

These disadvantages were recognized early, and many organizations have supported the concept of a database to control for many of the problems. The database is in effect a single super file containing all instances of all data, with a sophisticated management system to ensure that only legitimate users may access those parts of the data to which they have a need. Access can duplicate any structure the user is familiar with, but there is a major expense involved in maintaining and coordinating the database. The terms database management system and database administrator have been developed to reflect this cost.

As with files, three database structures have evolved. The hierarchial data structure consists of a series of root or parent nodes each pointing to several branches or descendents. By anticipating probable accesses, nodes with similar demand characterisitcs can be grouped to simplify access procedures. For example, if Part Type is the root structure, it is much easier to generate a list of all suppliers of a given part than it it is to find all parts from a given supplier. If both types of accesses are anticipated, the user can duplicate all information using the second structure (which is expensive) accept the penalty of time and inefficiency associated with the structure, or establish a second set of links between associated parts.

This later approach characterizes the network oriented database structure. Data is retained as discrete records with links connecting the various similar internal fields between records. Any record may have a large number of links coming in or going out (multiple roots/parents, multiple branches/children).

Unfortunately, both hierarchial and network structures cost the user substantial overhead, in many cases 200% - 300% of the space used to store the actual data.

The third solution to the problems associated with files that has become popular with the decline in cost of computational power is the relational database design. Records are stored as fixed relations in large tables. User requests are viewed as segments of the table, sometimes single fields, sometimes small sets of fields or records. Relational databases are generally considered superior to the othe designs because of the inherent flexibility in their design and their ability to respond to varying, unanticipated requests. Users of relational databases need to acquire some sophistication in their requests, as often requests may yield either a null set or substantially more information than is needed. The user specifies the characteristics to search the database for, and the exact matches are provided back to the user, much as a donut cutter may slice a donut from rolled out dough.

These traditional databases share some common strengths and weaknesses. Substantial overhead is needed to maintain links/structures of the database when compared to the traditional file structure. Their organization tends to favor numeric or short memo storage in contrast to lengthy text. Databases however, reduce waste due to redundant data, and can avoid (in centralized databases) or minimize (in distributed databases) inconsistencies within the data set. Data can be easily shared among various applications and users, while corporate, national, and industrial standards can be enforced. Any applicable security restrictions are easier to maintain in a database oriented environment while data integrity and accuracy can be maintained.

Springing from this file and database background, the hypertext concept has found fertile soil for the users of computer-based document systems. We shall now address the concept of hypertext.

The Concept of Hypertext/Hypermedia

Reading is fundamentaly linear. Words are grouped together to form sentences, sentences to form paragraphs, paragraphs to form documents. Each has a beginning and is read through to the end. If the reader is searching for a particular piece of information contained in the document, the document will be read (or skimmed) until the information is located, again in a linear fashion.

With large information oriented documents, such as texts, the reader may elect to use an index to locate the page containing the particular reference. The page is then read to locate the

information sought. In encyclopediae and dictionaries, the user can turn directly to the location of the information sought, but the act of extraction of the information will be linear. Obviously, the finer the division of indexing, or topic selection, the less reading is needed to extract the desired information.

While the analogy may be strained, one can view a linear document as corresponding to the sequential file structure, indexed documents as corresponding to the indexed sequential structure, and the encyclopedia/dictionary as a direct access structure. In all three however, the information sought still needs to be extracted from the surrounding document matrix through the linear task of reading.

Often relevant information about a topic will be contained at more than one location in a given document. Thus one encounters the multi-page references of indexes or the "see also" entries of the enclycopedia and dictionary. These cross references take on some of the characteristics of hierarchial and network data bases.

All of these referenced text products have been reproduced in computer usable formats. The ability of the computer to rapidly process file information and to search for particular addressable locations (as in a network oriented database) may permit the text developer an extra level ofreference ability, the ability to backward reference an information item. The user may consult an index (or menu) and bring up a particular piece of information. That particular piece of information may in turn lead to other locations in the data set. At any of these further locations, the user has the option of selecting one to many further references or of "looking the other way" to see from where the current location has been referenced. By establishing a view of both the incoming and outgoing references, the user can pursue data references in unique combinations of trails. Supported by proper hardware capabilities, this concept of multi-referenced nodes of information is termed hypertext. If illustrations, graphics, sound, etc. are added, the result is a hypermedia document.

One way to understand the importance of the referencing capabilities of hypertext is to momentarily digress and create a taxonomy of access methods to obtain information inprinted documents. At the 0th level we have no indexing, and searches for information are strictly sequential. At the 1st level, we place the index pointing directly to the topic area. Further complexity of indexing is seen in the encyclopedia or dictionary level in which the index points to topics which in turn point to other topics (see also...concept). This would be termed the 2d level. If we design our information retrieval system such that indexes

point to nodes which in turn point to other nodes (as in the 2d level) but the nodes carry information as to which nodes are pointing to them, so that the user may not merely backtrack but also explore alternate routings both forward and backward through the data set, we achieve the 3d level of indexing. By cross-linking our initial data set to reference footnotes, other data sets which might contain relevant information, further indexes, etc., we develop a truly comprehensive information retrieval system. Hypertext implementations are presently at this third and fourth level of functional development.

A well designed hypertext document is a fully referenced text organization. All items of potential interest are linked to explanations, related items, and other related text locations. For example, the text of poem which uses regional dialect or colloquialisms might provide a means to immediately substitute the modern expression for the regionalism (Guide demonstration disk), but might also be incorporated in turn into a data set of poems from the region, with the ability to isolate a geographic area and see a list of poetry (or whatever) associated with that area, perhaps by a graphic of a map and a freely moving cursor. In addition to poetry, expanded views of the area selected might also be optionally accessed, with historical and social commentary provided for background to the poetry.

Hypertextual organized reference materials promote fast access to the information desired. They also provide the user with a browsing capability of examining both the node sought and information about related nodes of information. This broadened picture very much strengthens the user's conceptual grasp of the data and the relations among the various nodes. Available alternatives become so rich that the problem arises of navigating through the nodes, and the concern arises that the user will become "disoriented" both as to how to continue on to their goal as well as how to return to their initial level of operation.

Hypertext Implementation Issues

As with so many other communication issues, the structure and organization of the nodes and links used in the hypertext document can reduce or heighten the disorientation problem. Proper sizing of nodes, recognition of the limits of human perceptual and cognitive limits, and other psychological and perceptual factors need to be remembered by the development team responsible for creating the document. Not all of these factors have been finally defined, particularly in regard to functional use in the hypertext environment.

In creating a hypertext document, information is collected in textual chunks sufficient to contain one idea or concept. Each of these nodes are linked to other nodes containing related material. While no absolute limit has been discussed in the literature, it has been observed by the author that more than seven links - either coming in or leaving the node - become difficult to manage for the user. Exceptions to this do occur, particularly where one is dealing with accessing a series of related cases, but for most uses this is a practical guideline.

One of the defining characteristics of a hypertext system is the presence of direct machine support for references between the nodes. This machine support implies that the user can jump from one node to another through single keystrokes or cursor movement/keystroke activities.

Hypertext concepts have been extended to several areas. The first of these is the on-line library or literary system. The documents in the library would be linked by machine supported hypertext nodes. Provision is made for users to add comments or criticisms and to respond to others' comments. Document creation and collaborative efforts can be supported through the underlying software systems. A current implimentation of this is available from McDonald-Douglas called NLS/Augment and is in active use with the Air Force.

A second category of hypertext implimentations are problem exploration tools. Because of the ability of hypertext to handle ideas and to quickly link those ideas, parallel concept generation is easily supported. Multiple ideas about a topic - or set of topics - can be created, with the author able to remain unconcerned about the relationships among the ideas during the initial creative stage. The linearity of thought required for traditional text generation can be bypassed during the initial creative rush. An example of these problem exploration tools is Maxthink from MaxThink Corporation.

The third category of hypertext systems are the browsing systems. These read-only systems are useful for teaching, reference and information systems. ZOG from Knowledge Systems, Inc. has been implimented as an intelligence review system for the Navy. The Interactive Encyclopedia (TIE) under development at the University of Maryland is a second example.

Most of the developed hypertext systems reviewed by the author fall into one of the three categories described above. Common to all is the presence of machine supported intertext references that permit the user to move forward and backward through the text. A mechanism for automatically marking the nodes or "trail" used through the document is also a common feature. Many permit

backward tracking to exit from a particular node trail, and most also permit a quick exit to some form of main menu or initial state to regain the "large picture". All but the purest of browsing systems allow the user to add annotations or links to the document to customize a general system to their individual needs.

Screen clutter becomes a problem as the number of nodes accessed grows. The user will often feel disoriented due to the many degrees of freedom available for movement at any point in a moderately complex hypertext document. To solve this different approaches have been implemented. Overlapping windows of text permit the user to see the most current text while leaving evidence of other windows present on the screen. Unfortunately, as the size of the current window is increased to accomodate increased text or to improve legibility, the evidence or presence of other windows disappears, requiring the user to backtrack until a cue to sigificantly earlier windows becomes visible.

Several implementations use an icon approach to maintain user awareness of alternate pathways. It can become a significant task for the user as well as for the developer to establish sufficiently unique icon designs to distinguish among many alternatives available, and to recognize the appropriate icon to select next. Both appropriate symbols and efficient abbreviations still require user training for maximum effective performance. Other approaches presently being studied involve rich graphic fields using rooms/doors/rooms analogies and flight analogies. Whether these approaches will prove effective given their relatively high cost in system overhead is still being examined.

Areas for research and related topics

A key to increasing the impact of hypertext on computer-based document systems in the future is the improvement of the user interface to take advantage of the innate strengths while minimizing the effects of the innate weaknesses present in the user. How should the presence or absence of further links or help alternatives be indicated? Is the use of special printing characters such as reverse video or special characters an appropriate technique or should icons be attached in some fashion to linked text areas? Should these be always visible or should they become visible with a keystroke (togglable)? How natural is the use of a hypertext designed document? Should naive users be expected to quickly find their way through it or will some training or orientation be required?

We know that users possess varying abilities of spatial

orientation and that these abilities can simplify the task of managing the data relationships present in a hypertext document. Users are able to retain several stages of nodal levels in their memory and are able to handle visual fields containing approximately thirty items before performance degrades significantly. Designers of hypertext documents may be able to work with these limits in the development of their work.

Areas of immediate importance to NASA for evaluating the utility of hypertext for space flight activities include a comparison between hypertext and traditional computer-based document storage systems. Specifically, training time to a given level of functionality, error rates during test trails, retraining needs after a significant non-rehersal time period, and graphic versus textual presentation approaches will need to be examined.

Summary

Hypertext represents a fascinating fieldfor exploration of the relationships between users, data, and computers. Work is proceding in using concepts from hypertext in fields as diverse as education and expert systems. Several commercial packages are being made available in the immediate future. It is my hope that the reader will continue to explore this subject area and pass along to the author any observations or products which should be found.

BIBLIOGRAPHY

Robert Akscyn and Donald McCracken, "ZOG and the USS CARL VINSON: Lessons in System Development",Carnegie-Mellon Technical Report No. CMU-CS-84-127, March 1984.

Stuart Card and Austin Henderson, "A Multiple, Virtual-Workspace Interface to Support User Task Switching", Intelligent Systems Laboratory, Xerox Palo Alto Research Center, Palo Alto, California 94304.

Jeff Conklin, "A Survey of Hypertext", MCC Technical Report No. STP-356-86, Rev. 1, February 1987.

Donna Erb, "Trip Report - May 18 - 22, 1987 JSC Missions Operations Training Task", MITRE Document

George Furnas, "Generalized Fisheye Views", Bell Communications Research, Morristown, New Jersey.

Guide User's Manual, Owl International, Inc., 144218 NE 21st Street, Bellvue, Washington,98007.

W.P. Jones and S. T. Dumais, "The Spatial Metaphor for User Interfaces: Experimental Tests of Reference by Location Versus Name",ACMTOOIS, Vol 4, No. 1, pp. 42-61, January 1986.

Neil Larson, MaxThink Journal, MaxThink, Piedmont, California, Vol. 2:3, January 1986.

Neil Larson, Houdini 2.1 Manual, MaxThink, Piedmont, California 94610

Jeffrey Young, "Hypermedia", Macworld, March 1987

MATHEMATICAL MODEL FOR THE DC-AC INVERTER FOR THE SPACE SHUTTLE

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1987

Johnson Space Center

Prepared by:	Frederick C. Berry
Academic Rank:	Instructor
University & Department:	Louisiana Tech Electrical Engr. Dep. Ruston, Louisiana 71270
NASA/JSC	
Directorate:	Engineering
Division:	Avionics Systems
Branch:	Avionics Integration
JSC Colleague:	Bob Hendrix
Date:	August 14,1987
Contract Number:	NGT 44-001-800

ABSTRACT

The purpose of this report is to inform the reader of what has been done for the mathematical modeling of the DC-AC inverter for the Space Shuttle. The mathematical modeling of the DC-AC inverter is an essential element in the modeling of the electrical power distribution system of the Space Shuttle. The electrical power distribution system which is present on the Space Shuttle is made up of three strings each having a fuel cell which provides DC to those systems which require DC, and the inverters which convert the DC to AC for those elements which require AC. The inverters are units which are two-wire structures for the main DC inputs and two-wire structures for the AC output. When three are connected together a four wire wye connection will result on the AC side. The method of modeling will be performed by using a Least Squares curve fitting method. A computer program will also be presented for implementation of the model along with graphs and tables to demonstrate the accuracy of the model.

INTRODUCTION

The purpose of this report is to inform the reader of what has been done for the mathematical modeling of the DC-AC inverter for the Space Shuttle. The mathematical modeling of the DC-AC inverter is an essential element in the completion of the modeling of the electrical power distribution system of the Space Shuttle which the in-house support contractor had started by producing a model for the DC portion of the electrical power distribution system. The electrical power distribution system which is present on the Space Shuttle is made up of three strings each having a fuel cell which provides DC to those systems which require DC, and the inverters which converts the DC to AC for those elements which require AC. The inverters are units which are two-wire structures for the main DC inputs and two-wire structures for the AC output. When three are connected together a four wire wye connection will result on the AC side. Therefore, with three strings on the Shuttle there are nine inverters. For this report only three inverters will be modeled. There was enough input data taken on the performance of the inverters that it was not necessary to write any equations describing the

internal operations of the inverter. Therefore the method of modeling will be performed by using a Least Squares curve fitting method. A computer program will also be presented for implementation of the model along with graphs and tables to demonstrate the accuracy of the model. The following table is given as an tool to help the reader identify any symbols that are used in this report.

V.....voltage
Hz.....frequency cycles per second
VA.....volt ampere
Pu.....per unit
Edc.....voltage DC
Idc.....current DC
Eac.....voltage AC
Iac.....current AC
PF.....power factor
Pin.....DC power in
Pout.....AC power out
EF & Eff.....efficiency
TF.....temperature

The first step taken in developing the math model for the inverter was to find out what function the inverter performs and what data was available on the inverter. The inverters are units which can convert 28VDC nominal input to a single phase AC output of 117V, 400Hz, rated at 750VA continuous, and 1125VA for 30 minutes with power factor ranging from 0.7 lagging to 0.9 leading including unity. The full rated load at 1.0Pu is:

$$I=750VA/117V$$

This defines the steady state operation of the inverter for constant inputs as a single phase unit.

Data on the inverter was available in the form of written reports and experimental data. The reports that were found and used are referenced at the end of this report. An example of the experimental data is given in Table 1. The experimental data shows the input and output conditions of the inverter keeping the power factor, input and output voltage, and frequency constant while changing the load.

%-VA	Edc	Idc	Eac	Iac	PF	Pout	TEMP	Pin	%-EF
0	30	3.14	117	0.00	.0	0.0	112.7	94.1	0.0
20	30	5.69	117	1.36	.6	96.8	111.6	170.5	56.8
40	30	8.63	117	2.59	.6	180.6	107.7	258.2	69.9
50	30	10.59	117	3.30	.6	232.9	108.1	316.6	73.3
60	30	12.16	117	3.88	.6	272.9	108.4	362.3	75.3
80	30	15.88	117	5.12	.6	361.6	110.8	473.7	76.3
100	30	20.39	117	6.47	.6	449.4	115.1	607.1	74.0
125	30	26.86	117	8.17	.6	579.5	122.6	797.4	72.6
150	30	33.53	117	10.05	.6	700.8	133.1	997.9	70.2
200	30	48.43	117	13.29	.6	926.8	123.8	1424.3	65.0

TABLE-1: EXAMPLE OF EXPERIMENTAL DATA FOR INVERTER A

This procedure was repeated for each of the three inverters for the following conditions: 24, 26, 28, and 30 volts DC input and -0.6, -0.7, -0.85, 1.0, and 0.9 power factor at each of the DC voltage inputs.

MODEL

The math model of the inverter was designed to produce as many input and output conditions as possible based on a limited number of inputs to the math model. The inputs to the model are as follows:

1. Percent Load
2. Input Voltage DC
3. Power Factor

the outputs from the math model are as follows:

1. Input Current DC
2. Output Current AC
3. Input Power DC
4. Output Power AC
5. Percent Efficiency
6. Temperature In Degrees F

In order to do this, two different modeling methods were investigated. These modeling methods are as follows:

1. Least Squares (LS) fitting of the data.
2. Data fitting by use of a set of orthognal polynomials (OP).

Each of these methods are found in software that is already available. The least squares fitting was found in the Statistical Analysis System (SAS) routines and the method of using orthogonal polynomials for curve fitting was found in the International Mathematical and

Statistical Libraries (IMSL) routines.

Based on analysis of the percent efficiency data it was decided that a fourth order model would best approximate the data. The reason for using the efficiency data to determine the order of the model was because this data went through the largest number of cycles over the range of 0 to 200 percent load. Using a fourth order model for each of the two cases listed above produced results which tracked the data very well. After being assured that each method produced satisfactory results, it was decided to use just the least squares fitting of the data to model the remaining sets of data. The reason for this decision was based on the observation that the least squares method of data fitting produced a tighter fit.

PROGRAM

Using the data that was supplied for each inverter a fourth order regression equation using the percent load as the independent variable was produced to model each of the six variables that were used from each sheet of the data for the inverters. Each inverter was tested at 24, 26, 28, and 30 volts DC input. At each voltage setting, it was tested at -0.6, -0.7, -0.85,

1.0, and 0.9 power factor. Using the results of these tests a set of regression equations was produced to represent the input current DC, input power DC, output current AC, output power AC, percent efficiency, and temperature in degrees fahrenheit for each setting of voltage and power factor. This produced a set of 120 regression equations for each inverter. Twenty regression equations are used to describe each of the above referenced output parameters.

Each set of 20 regression equations is used to produce a matrix for all voltage and power factor settings, for each inverter. An example of the matrices that represent all of the input paramaters for a load of 100 percent for inverter C is shown below:

	-0.6	-0.7	-0.85	1.0	0.9
24	24.02708	26.80123	33.31841	26.70104	40.88011
26	23.49872	27.47219	31.48396	37.36097	37.46921
28	20.90991	25.11707	28.68474	21.02460	35.56254
30	20.47581	23.31632	27.19581	33.18778	44.52682

DC INPUT CURRENT FOR 100 PERCENT LOAD FOR INVERTER C

	-0.6	-0.7	-0.85	1.0	0.9
24	64.79873	63.91091	64.56882	63.93892	63.57912
26	65.91304	64.58717	66.15194	63.23208	63.20073
28	64.79345	64.45176	65.94344	63.48076	63.68013
30	64.49525	65.50586	64.74864	58.59911	62.55650

AC OUTPUT CURRENT FOR 100 PERCENT LOAD FOR INVERTER C

	-0.6	-0.7	-0.85	1.0	0.9
24	562.0085	635.5890	784.0984	970.1697	960.8156
26	583.8451	673.9015	788.1099	962.5731	954.8917
28	587.7390	692.4550	790.6331	978.7694	970.1374
30	611.5838	690.3112	815.6928	1022.1897	1008.9542

DC INPUT POWER FOR 100 PERCENT LOAD FOR INVERTER C

	-0.6	-0.7	-0.85	1.0	0.9
24	457.3632	522.5542	639.8809	746.2821	669.2880
26	461.2574	533.8483	636.5393	738.5239	666.9812
28	450.8919	528.9160	643.1387	741.7915	672.3812
30	454.0515	539.8164	643.8911	746.7502	670.5568

AC OUTPUT POWER FOR 100 PERCENT LOAD FOR INVERTER C

	-0.6	-0.7	-0.85	1.0	0.9
24	76.98697	80.62087	79.11906	74.81590	67.30208
26	76.46641	77.63451	83.97259	74.75730	68.18011
28	76.08212	76.09306	79.57481	74.34234	67.70587
30	74.93172	71.25590	77.79884	71.92826	64.74466

PERCENT EFFICIENCY FOR 100 PERCENT LOAD FOR INVERTER C

	-0.6	-0.7	-0.85	1.0	0.9
24	10.61268	11.22281	11.32683	11.76665	13.42805
26	11.28422	12.14870	11.54441	12.57221	12.17178
28	10.99731	11.23885	12.34068	10.67528	12.72728
30	11.89918	12.72210	12.58744	13.61134	13.33829

TEMPERATURE FOR 100 PERCENT LOAD FOR INVERTER C

This method of reproducing the data was chosen over having the data stored in a data file and reading the desired portion of the data file into the program.

At this point the program will perform a least squares on each matrix for each voltage setting using the power factor as the independent variable. This will produce four regression equations which will then be used to calculate four points, one for each voltage setting. With these four points another least squares will be performed using the voltage as the independent variable. This will produce one equation which will then give the desired value of current for the input of the percent load, input voltage DC, and power factor.

This procedure was repeated for each of the three inverters. Once this point is reached the following output is produced:

	INVERTER-A	INVERTER-B	INVERTER-C
Input Current DC..	27.19581	27.05088	27.21824
Output Current AC.	6.47486	6.27590	6.35214
Input Power DC....	815.69282	803.43621	808.26576
Output Power AC...	643.89115	624.95208	631.83659
Inverter % Eff....	77.79884	76.52319	75.55965
Inverter Temp. F..	125.87446	117.18629	107.85057

OUTPUT FOR EACH INVERTER AS A SINGLE PHASE UNIT

This represents the output for three single phase inverters. The next step is to combine these three inverters into a 4 wire wye connection with the

following phase sequence: inverter A, 0 degrees; inverter B, 120 degrees; and inverter C, 240 degrees. Using this phase sequence the next set of output that is supplied is the voltage line to neutral, phase current, neutral current, and power. This output is now represented as follows:

	REAL	IMAG.
Inverter-A Output Voltage	117.00000	0.00000
Inverter-B Output Voltage	- 58.49374	101.32858
Inverter-C Output Voltage	- 58.51251	-101.31774
Total Three Phase Voltage	- 0.00625	0.01084
Inverter-A Output Current	4.58541	4.67805
Inverter-B Output Current	- 6.06423	1.01328
Inverter-C Output Current	1.71725	- 6.16466
Total Return Current	0.23842	0.07286
Inverter-A Output Power	377.87152	385.50606
Inverter-B Output Power	361.95492	369.26788
Inverter-C Output Power	367.28991	374.71066
Total Three Phase Power	1107.11637	1129.48461

THREE PHASE OUTPUT FOR THE INVERTERS

A complete set of graphs is provided in appendix A for inverter A which shows the results of the math model calculated values against the experimental values for the following set of pramaters:

	EXPERIMENTAL	CALCULATED
Input Current DC	IDC	IDCC
Output Current AC	IAC	IACC
Power In	PIN	PINC
Power Out	POUT	POUTC
Percent Efficiency	EF	EFC
Temperature	TF	TFC

CONCLUSION

Thus far three inverters have been modeled by using the input/output data that was supplied. This data was modeled by use of a Least Squares fitting technique to produce a set of equations which approximates the output conditions and also some of the input conditions, given certain inputs to the model like percent load, input voltage DC, and power factor.

At this point only a steady state model for a balanced three phase system exists for one set of inverters. It is proposed to complete the steady state modeling by performing the following tasks:

1. To increase the accuracy of the model by increasing the order of the system or by use of other modeling methods.
2. To give the model the ability to handle unbalanced loads.
3. To include into the model an option that would allow for dropping of an inverter and then calculating the output of the remaining inverters.

With these additions to the present model a very good representation of the DC-AC Inverter for the steady state can be produced.

REFERENCES

IMPLEMENTATION AND OPERATION PROCEDURES
Prepared by Lockheed Engineering, Houston, Texas
Contract NAS9-15800

SPECIFICATION-INVERTER, POWER STATIC, 117 VOLT, SINGLE PHASE, 400 Hz
Prepared by Rockwell International, Downey, California

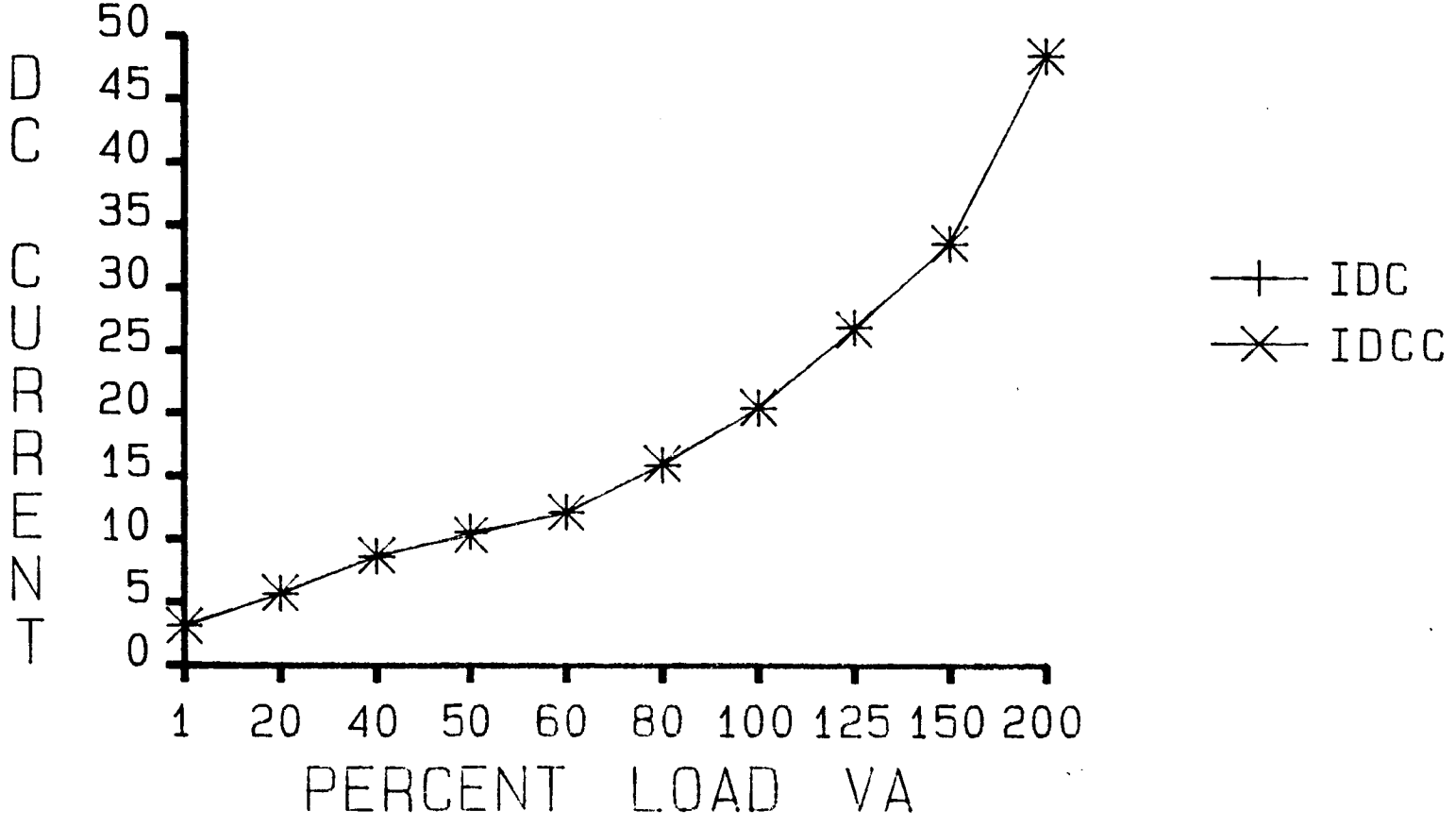
QUALIFICATION TEST PROCEDURE SPACE SHUTTLE POWER STATIC INVERTER, 117 VOLTS, SINGLE PHASE, 400Hz
Prepared by Westinghouse, Lima, Ohio
Document N5-118

FINAL TEST REPORT FOR 0V-101 EPDC BREADBOARD EVALUATION TEST SERIES
Internal Note JSC 77-EG-12

INTERIM TEST REPORT FOR 0V-102 BREADBOARD EVALUATION TEST SERIES
Internal Note ASD 78-EH-16

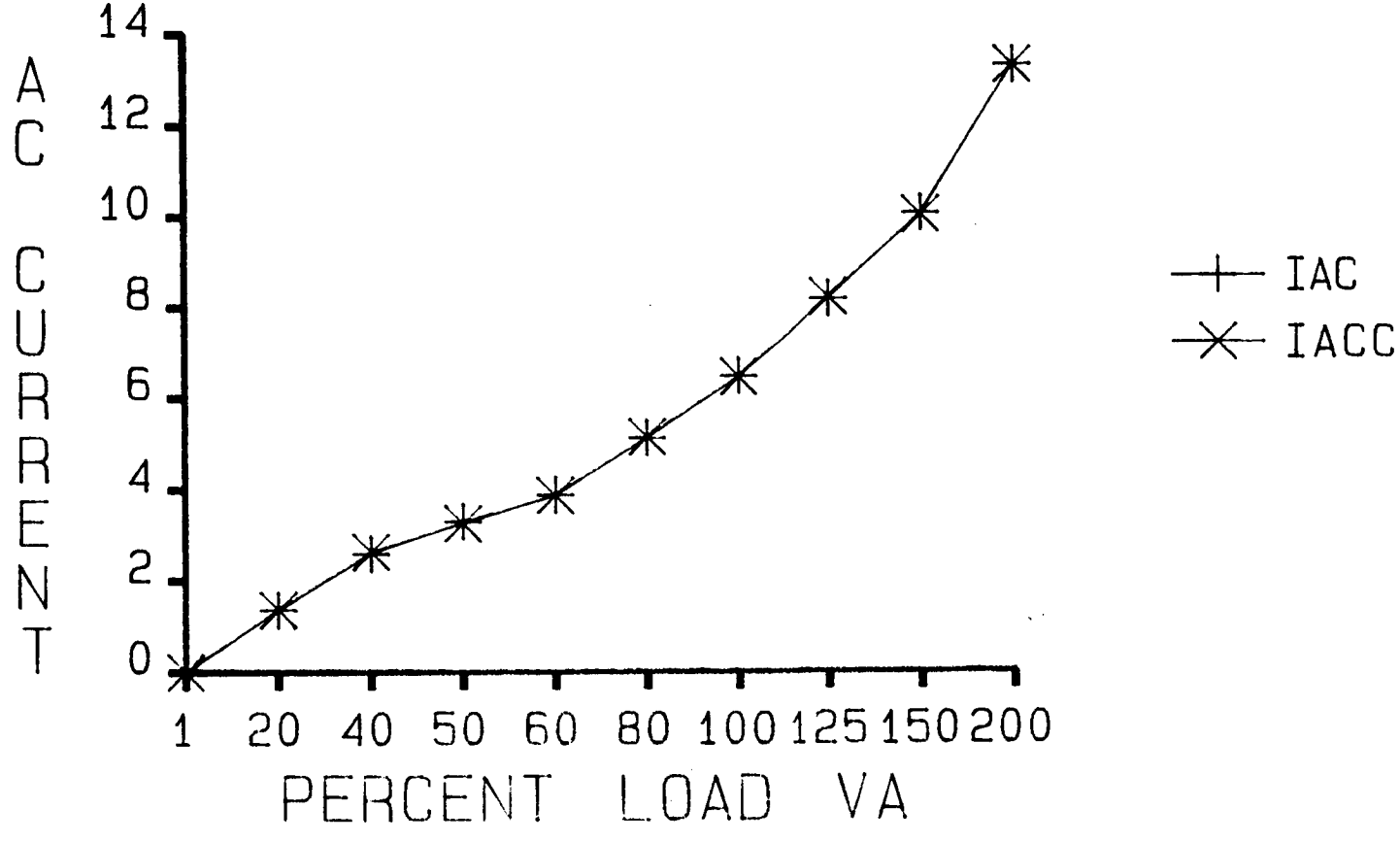
INVERTER A TEST RESULTS

FOR -0.6 P.F. AND 30 VDC



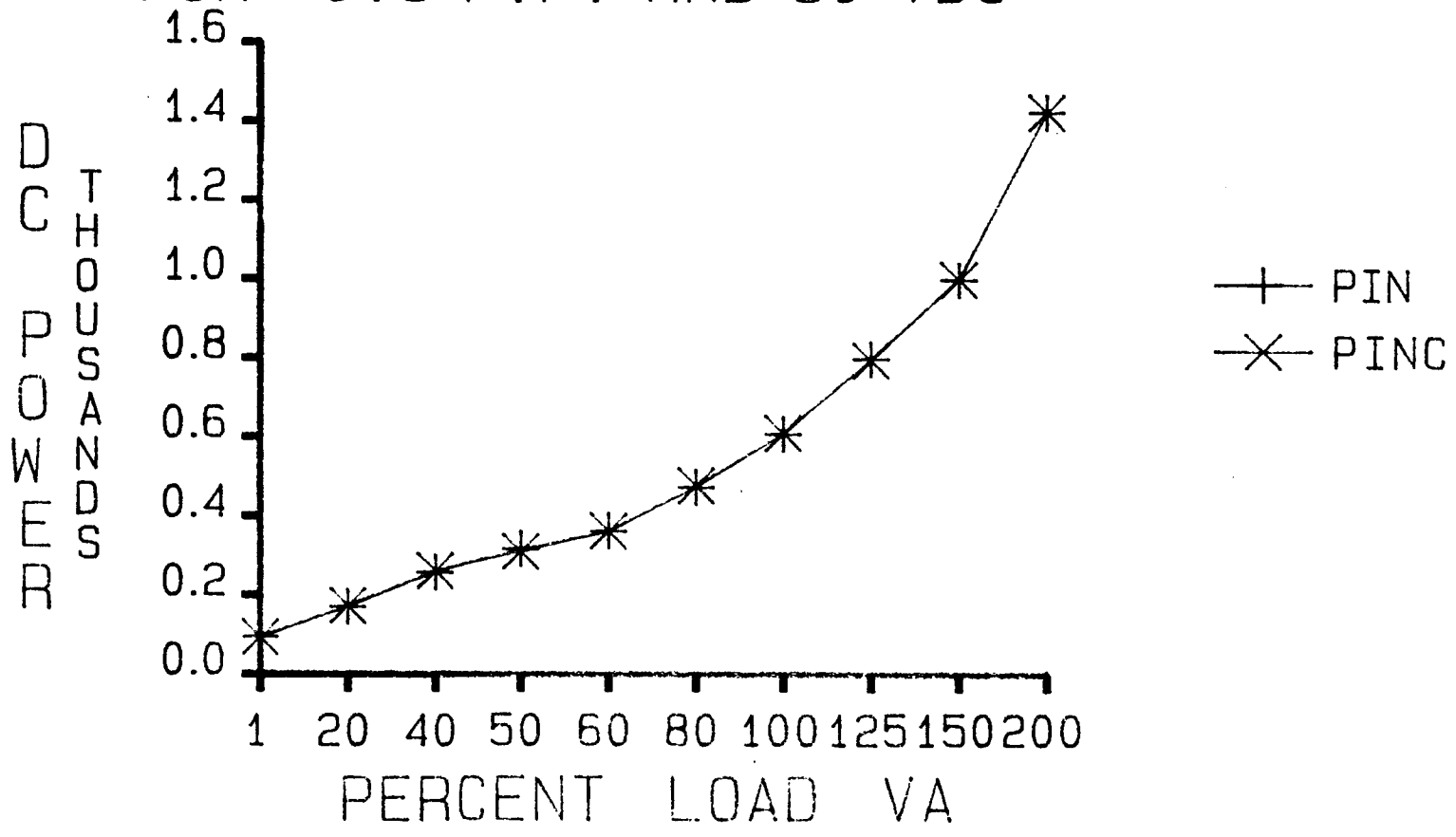
INVERTER A TEST RESULTS

FOR -0.6 P.F. AND 30 VDC



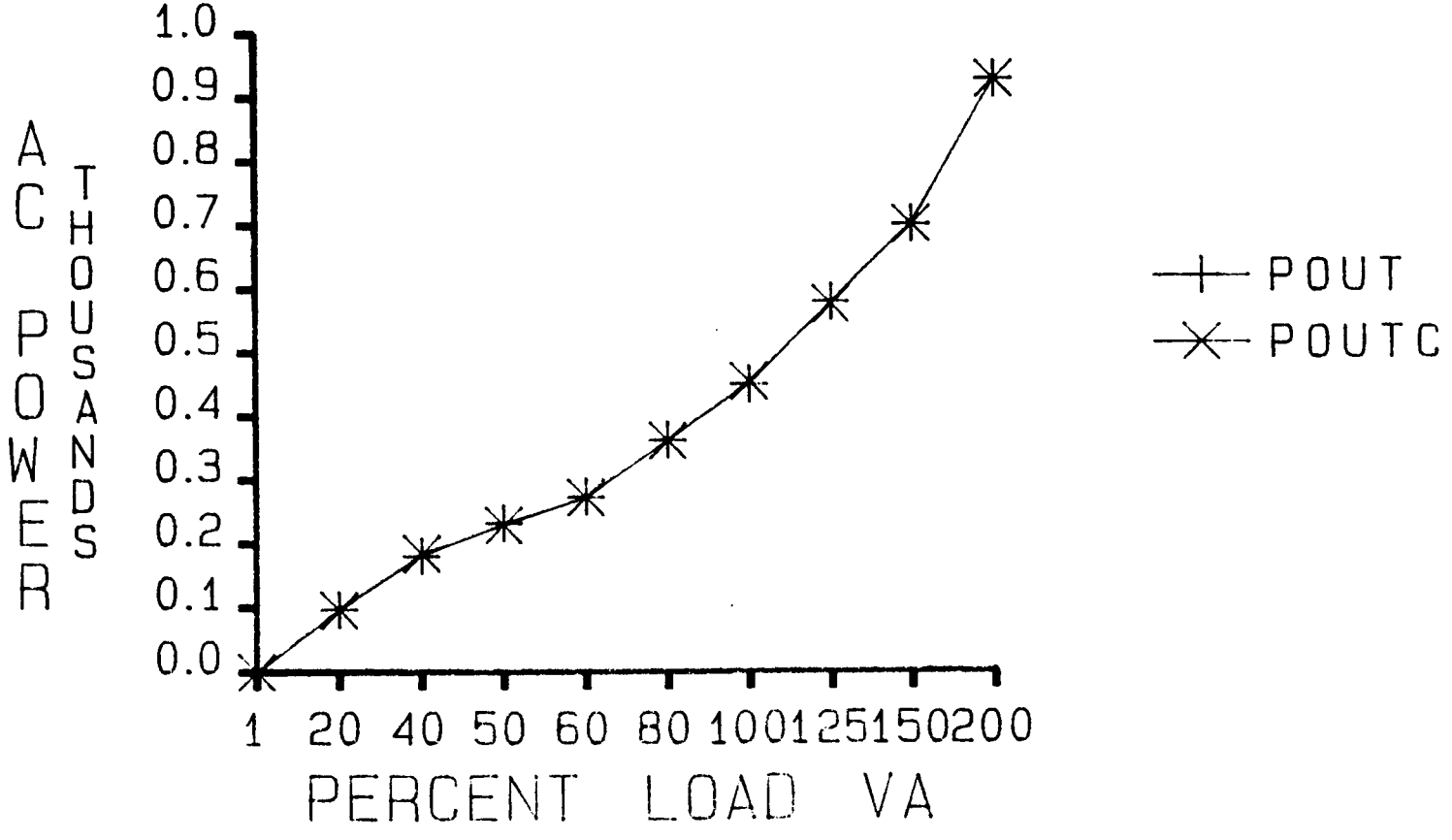
INVERTER A TEST RESULTS

FOR -0.6 P.F. AND 30 VDC



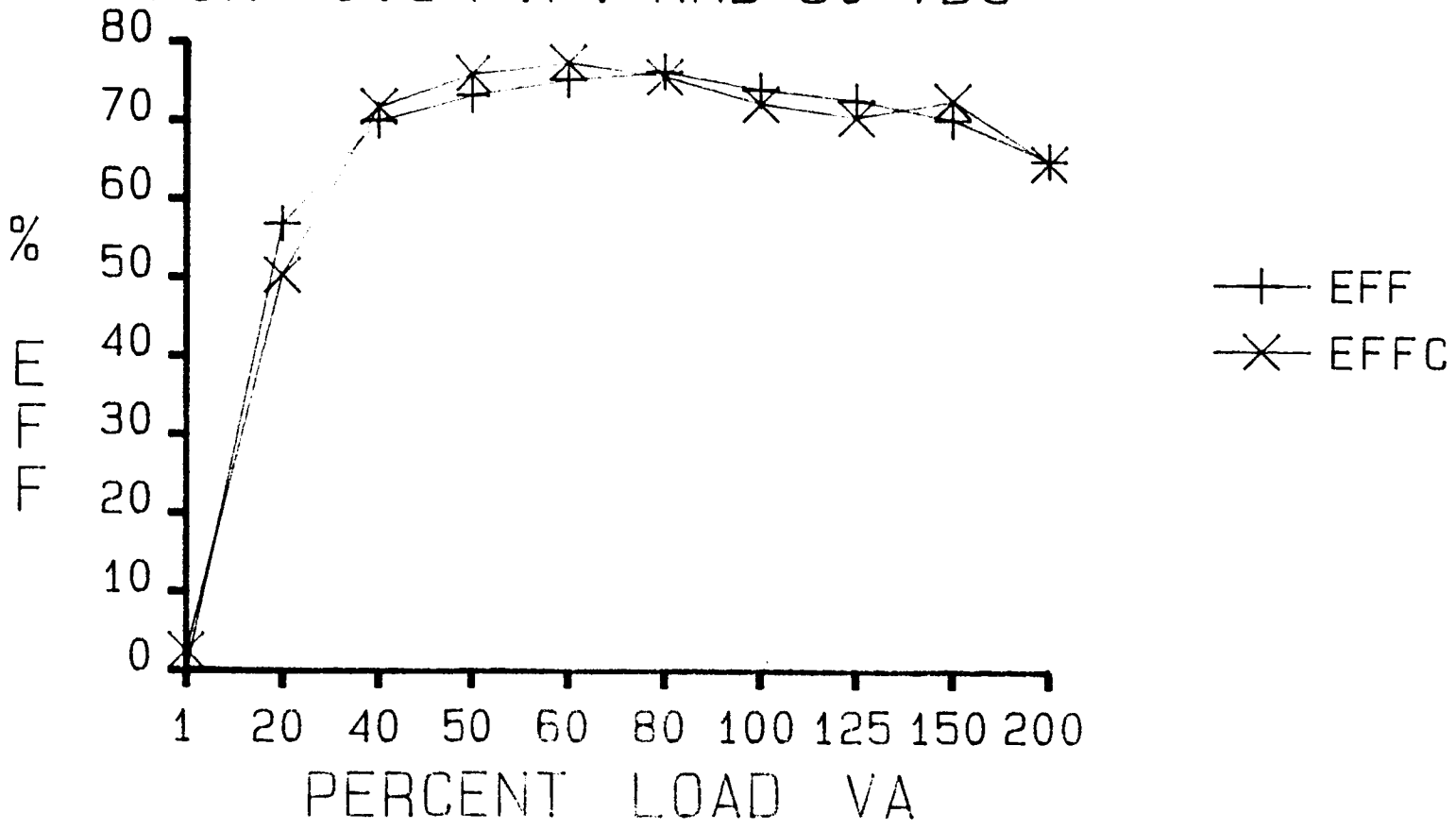
INVERTER A TEST RESULTS

FOR -0.6 P.F. AND 30 VDC



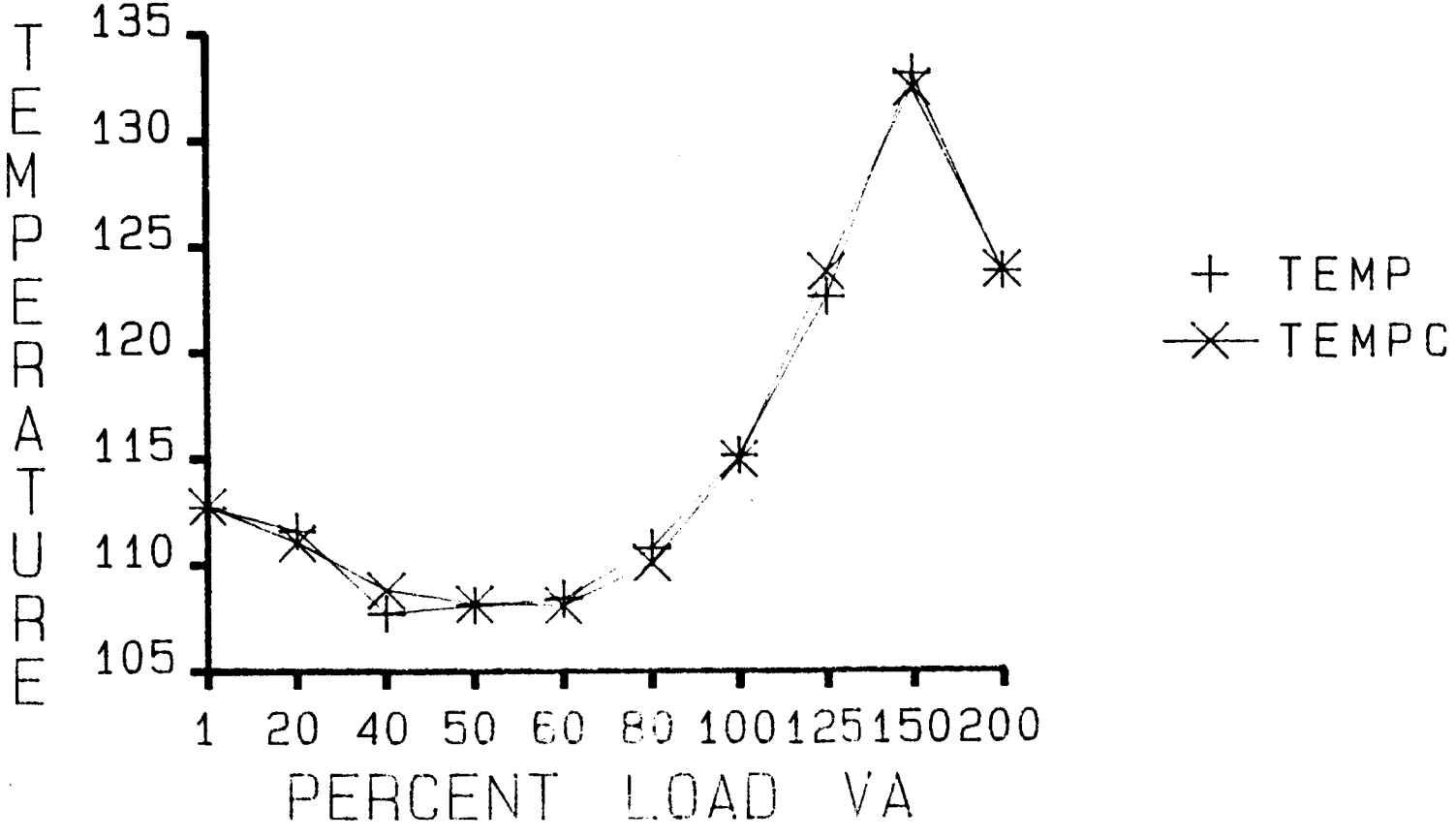
INVERTER A TEST RESULTS

FOR -0.6 P.F. AND 30 VDC



INVERTER A TEST RESULTS

FOR -0.6 P.F. AND 30 VDC



ANALYSIS OF LASER EXTRACTED VOLATILES IN CARBONACEOUS CHONDRITES

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1987

Johnson Space Center

Prepared by:	George E. Blanford, Ph.D.
Academic Rank:	Professor
University & Department:	University of Houston - Clear Lake Programs in the Natural Sciences Houston, TX 77058
NASA/JSC	
Directorate:	Space and Life Sciences
Division:	Solar System Exploration
Branch:	Experimental Planetology
JSC Colleague:	Everett K. Gibson, Ph.D.
Date:	August 11, 1987
Contract Number:	NGT 44-001-800

ABSTRACT

It is scientifically important to understand the composition of volatile compounds from interplanetary dust particles (IDP's) because they may be related to the primordial inventory of planetary materials which were necessary to provide environments conducive to the formation of life. The principal task for this project was to evaluate the use of a laser microprobe to measure volatiles in IDP's. Because primitive meteorites are thought to be closely related to IDP's, carbonaceous chondrites were used for the evaluation. Three sets of experiments were performed to determine the volatiles released from potential substrate materials, to analyze the volatiles released from matrices of bulk samples of carbonaceous chondrites, and to analyze volatiles released from ~100-200 μm meteorite particles to simulate IDP's. Aluminum appeared to be the best choice of substrate material. Mass ratios between carbonaceous chondrite matrices of Allende and Murchison show fair reproducibility with somewhat high uncertainties. Particles (~100 - 200 μm) from the Orgueil, Murchison, and Allende meteorites produced measurable quantities of volatiles that appear to have mass spectra comparable to the bulk matrices.

INTRODUCTION

We used a laser microprobe to study volatile compounds released from meteorites. Our ultimate view is to apply the laser microprobe to the study of interplanetary dust particles (IDP's) (Gibson and Sommer, 1986). These particles are so small (~10^{-6}-10^{12}g) that the laser microprobe may be one of the few practical ways to study volatile compounds in them. These compounds probably played an important role in establishing life in the solar system when similar materials were accreted to form the Earth or were captured after the Earth cooled. IDP's may be similar to primitive classes of meteorites and, therefore, meteorite studies provide useful preliminary work for establishing quantitative methods and for determining limits of reproducibility.

Laser Microprobe

The laser microprobe (Sommer and Gibson, 1985, 1986; Carr and Gibson, 1987) consists of a sample chamber mounted on a microscope stage which has a mass spectrometer in its evacuation line. The sample can be viewed and aligned using microscope optics. A mirror permits focusing a laser beam on the same spot. The laser is a Jarrell-Ash Q-switched Nd-glass laser with energy input to the sample of 0.1-1.0 J. When the laser is fired heat generated by the beam volatilizes part of the sample. The energy mode that was used for most of the analyses reported here (~0.56 J) created a heat-damaged crater of ~200-300 μm diameter in most samples. The gases released passed through a Hewlett-Packard model 5970 Mass Selective Detector (MSD). In this instrument the the gases are ionized, selected by a quadrupole mass spectrometer, and

detected by an electron multiplier. The MSD is controlled automatically by a Hewlett-Packard model 9000 series 300 microcomputer using Hewlett-Packard ChemStation software. This instrument and software are more conventionally used with a gas chromatograph replacing the laser excited sample chamber.

Interplanetary Dust Particles

Dust particles in the solar system are a valuable resource for planetary materials. These particles can be collected on spacecraft and high altitude aircraft, found in sea sediments, cryoconite deposits in Greenland, and in the ice sheet of Antarctica. Currently the curated particles (Zolensky et al., 1986) collected in the stratosphere provide the best source of documented particles many of which are IDP's.

Interplanetary dust particles must be selected from the many particles collected in the stratosphere. We attempt to do this by finding similarities in the composition and structure of the particles with known extraterrestrial materials. Many IDP's seem to have compositions closely related to meteorites, especially the more primitive classes of meteorites (Brownlee et al., 1987; Blanford et al., 1987). Recently direct measurements of cometary dust have been made and some IDP's seem to have compositions related to cometary dust as well (Kissel et al., 1986a, b; Balsiger et al., 1986). The most interesting fact is that, although compositions are closely related to other solar system materials, IDP compositions show their own uniqueness. Some of the IDP's are rich in carbon which may be present in organic species (Fraundorf, 1981; Blanford et al., 1987). These organic species are particularly important because of their possible relation to the formation of life in the solar system. Increased knowledge of the primordial organic inventory of the solar system will probably provide strong constraints on life forming events.

EXPERIMENTAL WORK

To obtain useful information from IDP's it will be necessary to analyze small particles quantitatively and reproducibly. Three sets of experiments were performed to determine the volatiles released from potential substrate materials, to analyze the volatiles released from matrices of bulk samples of carbonaceous chondrites, and to analyze volatiles released from ~100-200 μm meteorite particles to simulate IDP's.

Substrate Selection

The laser beam will necessarily be larger than particles analyzed. Consequently the particles must be placed on substrates which, when cleaned, contribute either an insignificant amount of released volatiles when compared to the intended samples or a measurable background of volatiles that can be subtracted from unknowns. The following materials were examined: aluminum, beryllium, indium, tungsten, tantalum, alumina, fused silica, silica air gel, borosilicate glass, metallized Kapton®, and Torr Seal®. The metals showed the lowest backgrounds consisting essentially of adsorbed air, but the metals themselves vaporized and condensed on the chamber window. Cleaning materials, namely detergents, isopropanol, and Freon®, did not give significant mass peaks. Aluminum appeared to be the best choice for present because the vaporized metal could be easily removed at the end of a sample run by chemical dissolution. A sample of polished aluminum did not give any improvement over conventional sheet aluminum. Total background signals from aluminum are as high as 25% of the signal from the small Allende grains to be discussed below, but become less significant to insignificant for samples with high volatile contents. Figure 1 is a mass spectrum from a single laser shot on cleaned aluminum which can be compared to other mass spectra of meteorites to be presented.

Stratospherically collected IDP's are entrapped on a plate covered with silicone oil which is subsequently dissolved with hexane (Zolensky et al., 1986). We measured laser

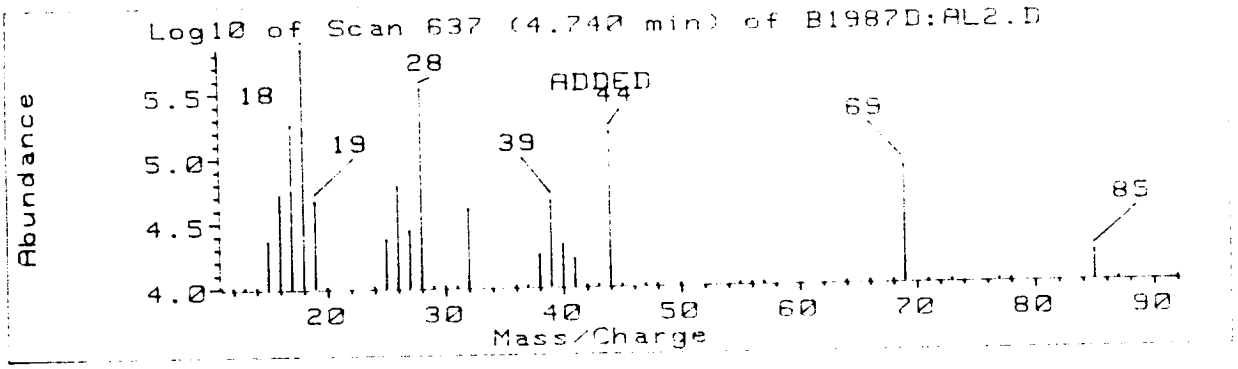


Figure 1. Mass spectrum of released volatiles from a single laser shot on cleaned aluminum.

extracted volatiles from silicone oil (on aluminum) and from a similar sample disk cleaned with hexane. Silicone oil produces a substantial amount of vaporized gas (Fig. 2) although a sample disk cleaned with hexane (Fig. 3) appears little different from aluminum cleaned normally (Fig. 1). We intend to continue to monitor for the presence of silicone oil and hexane to determine whether the collection and curation procedures disturb the laser extracted volatiles that will be measured from IDP's.

Meteorite Matrices

Laser extracted volatiles were measured from the matrices of bulk samples of the Murchison C2 and the Allende C3 carbonaceous chondrites to determine the amount and

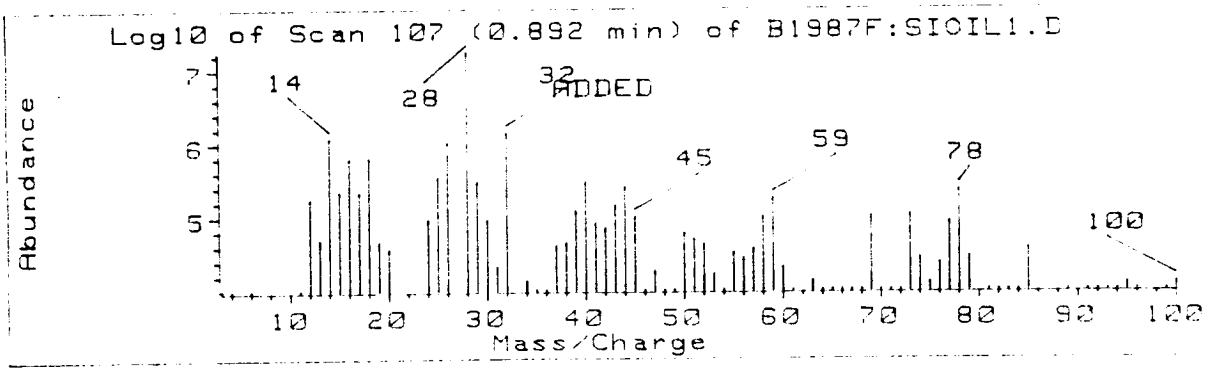


Figure 2. Mass spectrum of released volatiles from a single laser shot on aluminum covered with silicone oil.

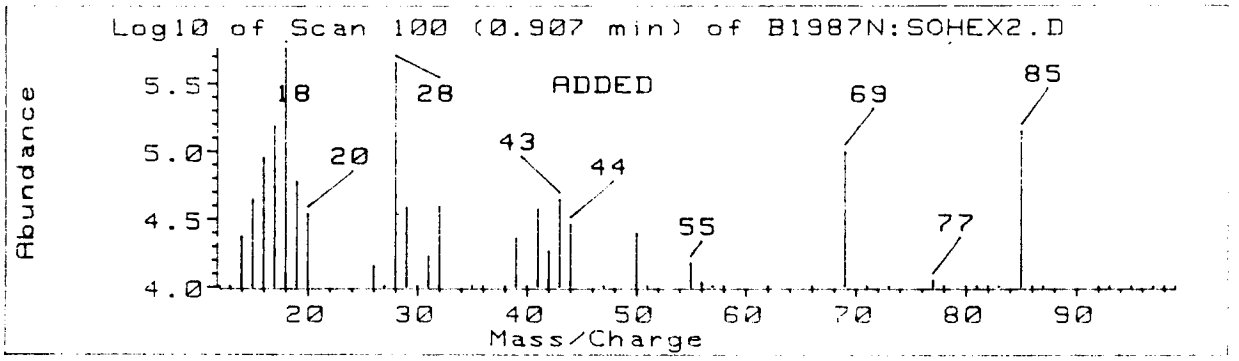


Figure 3. Mass spectrum of released volatiles from a single laser shot on aluminum, previously covered with silicone oil and then cleaned with hexane.

spectrum of gases that could be obtained from a fixed laser energy (~0.56 J). Data have been gathered for 20 spots in each meteorite from bulk samples each having ~1 cm^2 cross-section. Figures 4 and 5 show mass spectra from a single laser shot for Murchison and Allende respectively. It is difficult to compare the spectra from these graphs, however, because the volatile yield (counts) from Allende is lower than that for Murchison. Complete analysis of the data awaits the connection of the Hewlett-Packard computer to a Digital Equipment Corporation VAX® computer to statistically analyze the output. Because this connection has not been made, we must report incomplete analyses. Figure 6 shows averaged mass spectra for five laser shots on Murchison that

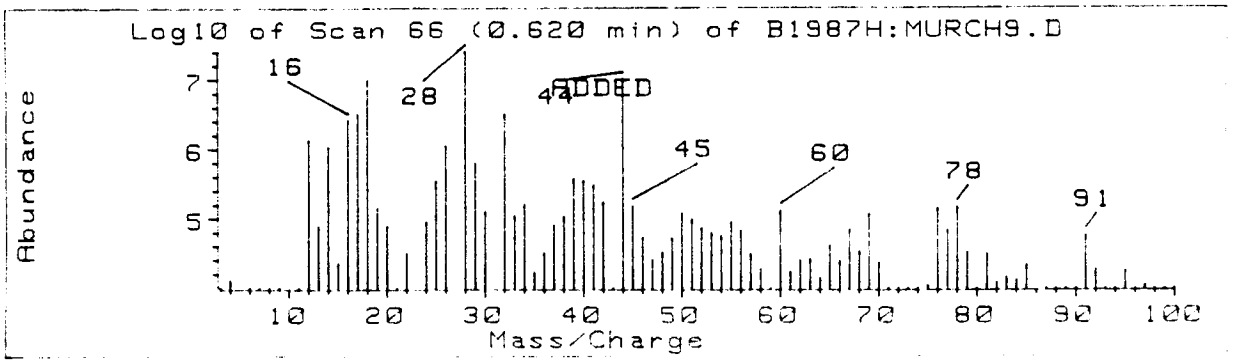


Figure 4. Mass spectrum of released volatiles from a single laser shot on the matrix of the Murchison meteorite.

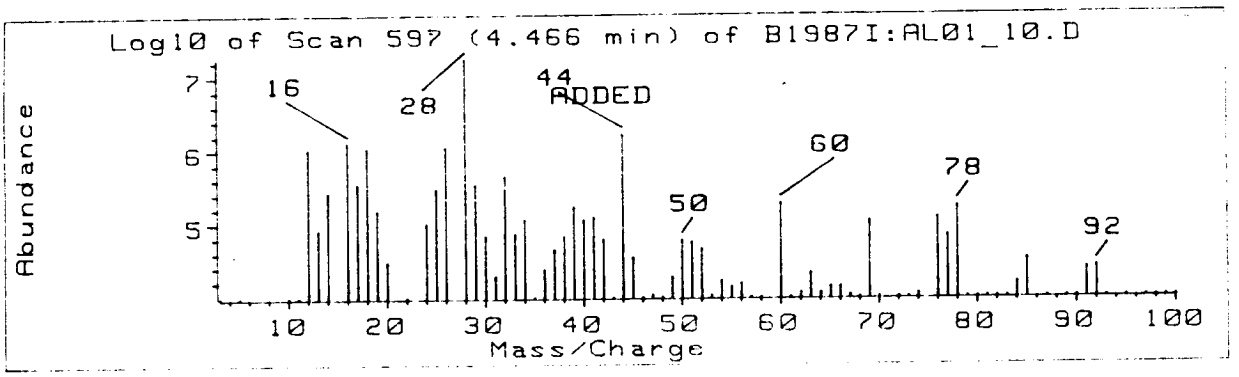


Figure 5. Mass spectrum of released volatiles from a single laser shot on the matrix of the Allende meteorite.

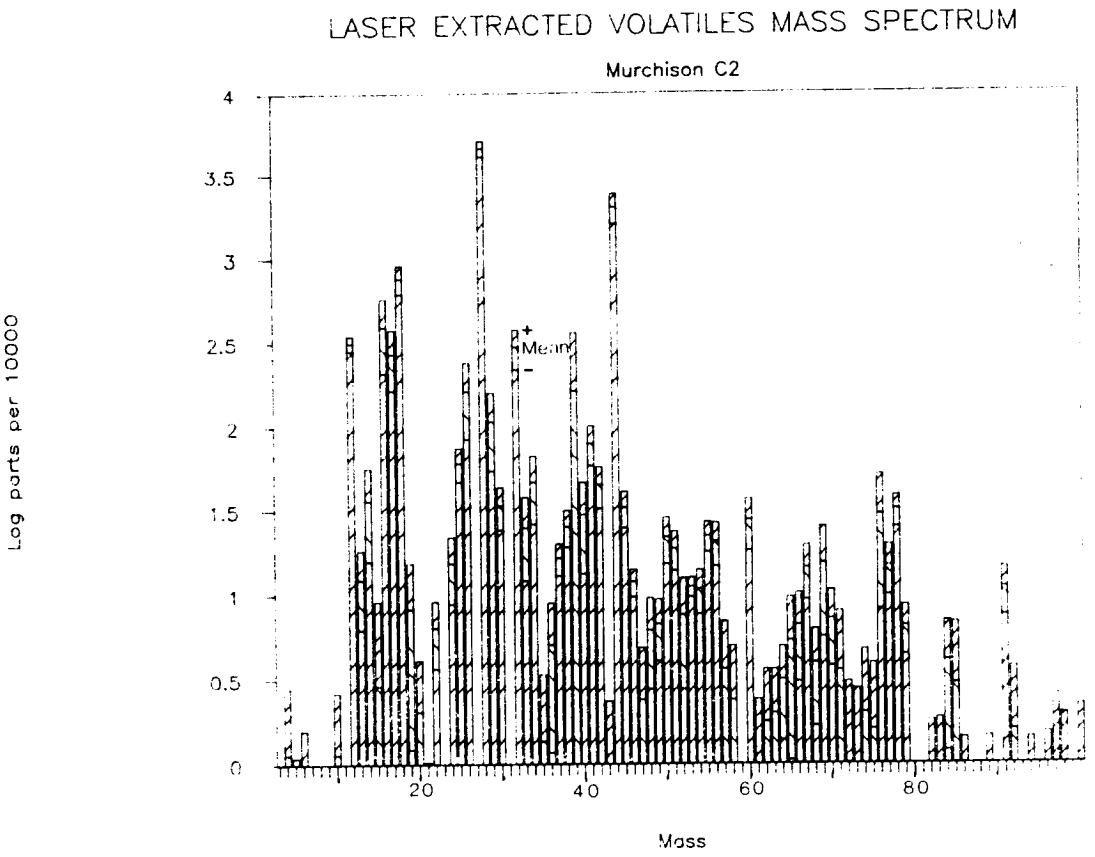


Figure 6. Average mass spectrum of released volatiles from five laser shots on the matrix of the Murchison meteorite showing the standard deviation for each mass.

C-2

was obtained by transferring data by hand. The data are expressed as parts per ten thousand of the total volatile signal and the mean and standard deviations are shown for each mass peak. Determining the standard deviation for each mass will enable us to find the reproducibility of successive analyses and the noise level for each mass. Table 1 shows a comparison of several selected masses and the total signal from a small subset of the Murchison and Allende bulk data. As expected the Allende meteorite yields a smaller volatile component than the more primitive Murchison meteorite. However, this initial look at the reproducibility of the data show uncertainties which seem large.

Meteorite Grains

Our final experimental data has been taken from 100-200 μm particles of the Orgueil C1, Murchison C2, and Allende C3 carbonaceous chondrites. The particles are larger than we expect for IDP's, but we will work with a set of smaller particles after evaluating the current experimental data. Again we need the computer link to complete the analyses of these data. Figures 7, 8, and 9 show mass spectra of a single particle from Orgueil, Murchison, and Allende respectively. Orgueil and Murchison seem to have similar yields which are higher than that of the less primitive Allende meteorite.

Table 1

Allende/Murchison

Mass	Signal	Fraction
18	0.032 ± 0.024	0.075 ± 0.068
28	0.632 ± 0.272	1.388 ± 0.263
44	0.272 ± 0.137	0.642 ± 0.200
Total	0.444 ± 0.202	

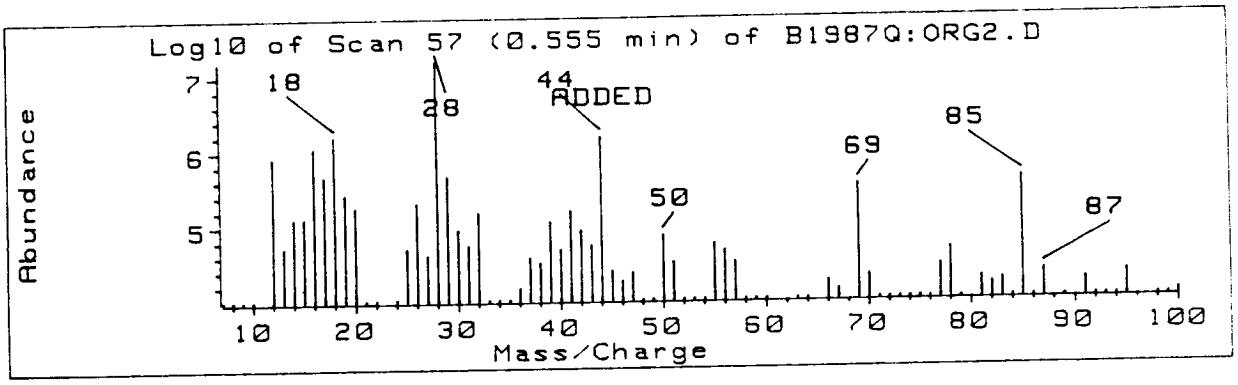


Figure 7. Mass spectrum of released volatiles from a 200 μm particle of the Orgueil meteorite.

The mass spectra of several grains of the Murchison meteorite have been averaged together in Fig. 10 which can be compared to Fig. 6 for the bulk meteorite. Table 2 compares some of the mass peaks obtained from the bulk and from the grains.

CONCLUSION

Considerable experimental data has been collected from carbonaceous chondrites to quantify and to test reproducibility of a laser microprobe. Until the data analyses are complete it is premature to examine the details of the mass spectra. It is clear, however, from the meteoritic mass spectra shown in this report that the laser microprobe releases fairly reproducible amounts of high mass volatiles from these meteorites both in bulk and as particles. However, the laser vaporized gases may react with themselves in

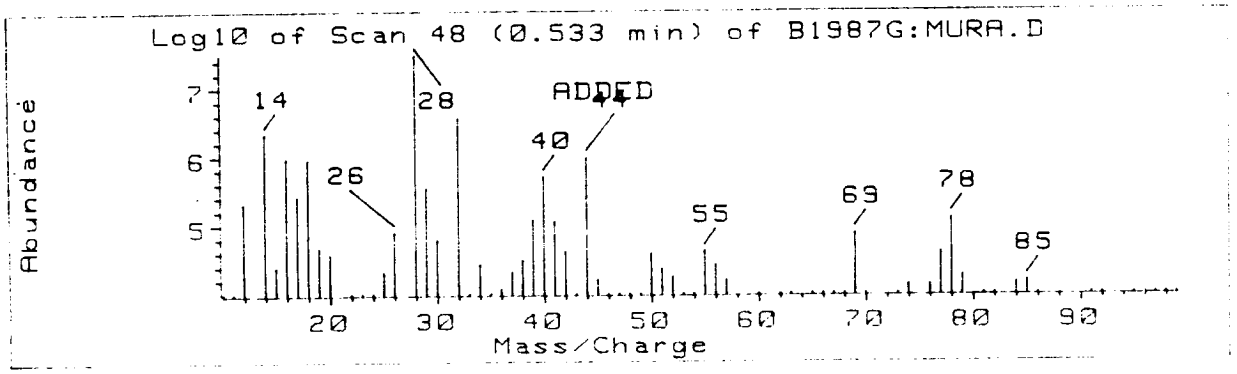


Figure 8. Mass spectrum of released volatiles from a 150 μm particle from the Murchison meteorite

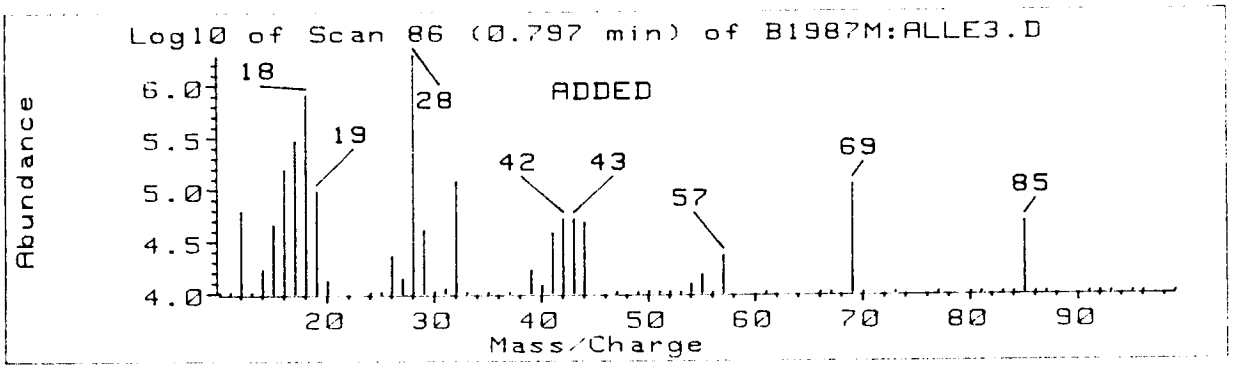


Figure 9. Mass spectrum of released volatiles from a 100 μm particle from the Allende meteorite.

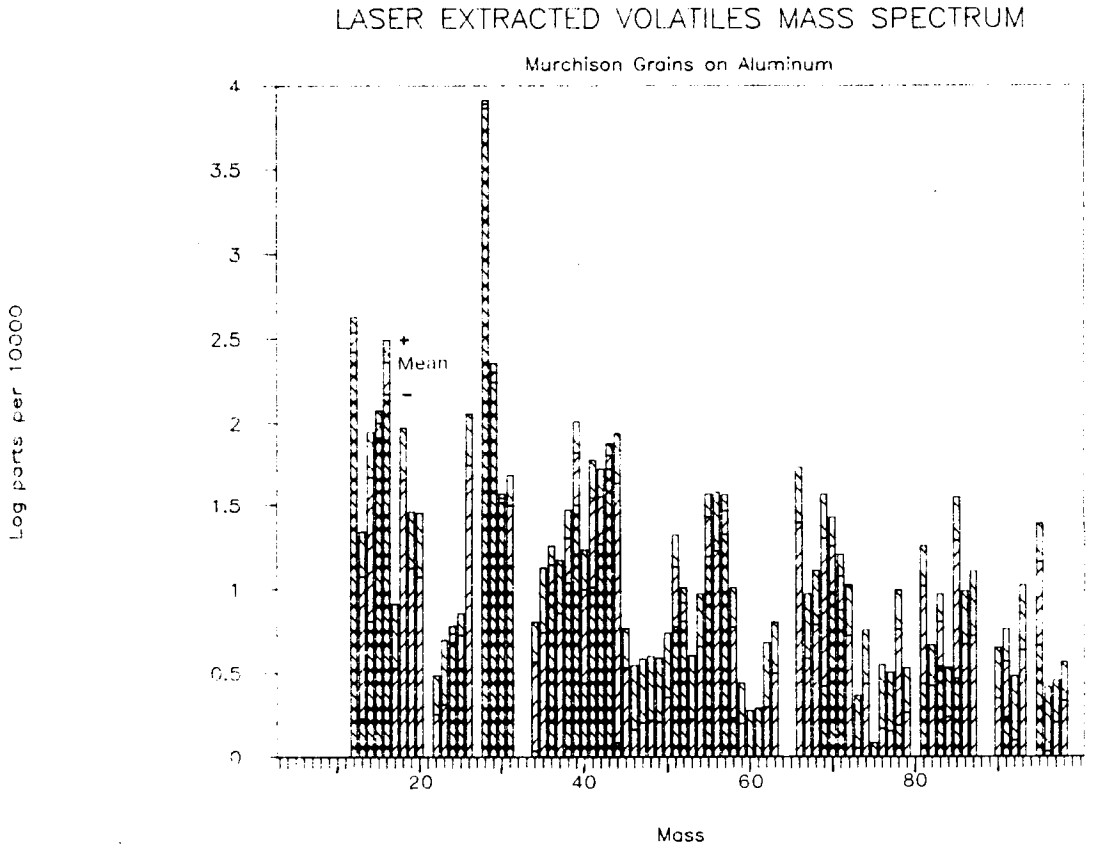


Figure 10. Average mass spectrum of released volatiles from three 100-200 μm particles from the Murchison meteorite showing the standard deviation for each mass.

Table 2

Selected Masses from Murchison C2

(parts per 10 000)

Mass	Particles	Bulk
12	344 ± 80	315 ± 32
13	12 ± 10	12 ± 6
14	48 ± 41	36 ± 20
15	100 ± 18	3 ± 6
16	229 ± 81	392 ± 183
17	3 ± 5	271 ± 106
18	46 ± 48	765 ± 148
24	5 ± 1	15 ± 7
25	5 ± 2	61 ± 14
26	56 ± 57	162 ± 78
28	7757 ± 433	4522 ± 475
29	200 ± 26	107 ± 52
30	35 ± 2	34 ± 10
32	0 ± 0	296 ± 78
44	44 ± 43	2005 ± 406
69	20 ± 17	15 ± 10
91	4 ± 2	11 ± 3
95	15 ± 10	2 ± 2

unknown ways and the ionizer may break them up further so that the indigenous volatiles in the meteorites are not immediately evident from the mass spectra. Interpretation of the mass spectra will require additional study. Almost certainly, however, a database of mass spectra from 38 000 compounds that is part of the ChemStation software will be useful in this interpretation.

REFERENCES

Balsiger H., Altwegg K., Bühler F., Geiss J., Ghielmetti A.G., Goldstein B.E., Goldstein R., Huntress W.T., Ip W.-H., Lazarus A.J., Meier A., Neugebauer M., Rettenmund U., Rosenbauer H., Schwenn R., Sharp R.D., Shelley E.G., Ungstrup E., and Young D.T. (1986) Ion composition and dynamics at comet Halley, Nature, 321, 330-334.

Blanford G.E., Thomas VerPloeg K., McKay D.S. (1987) Microbeam analysis of interplanetary dust particles for major elements, carbon and oxygen. Submitted to Proc. Lunar Planet. Sci. Conf. 18.

Brownlee D.E., Wheelock M.M., Temple S., Bradley J.P., and Kissel J. (1987) A quantitative comparison of comet Halley and carbonaceous chondrites at the submicron level. In Lunar and Planetary Science XVIII, pp. 133-134. Lunar and Planetary Institute, Houston.

Carr R.H., and Gibson E.K. (1987) A quadrupole mass spectrometer - laser microprobe system for the analysis of gases released from geologic samples. In preparation.

Fraundorf P. (1981) Interplanetary dust in the transmission electron microscope: diverse materials from the early solar system, Geochim. Cosmochim. Acta, 45, 915-943.

Gibson E.K., and Sommer M.S. (1986) Laser microprobe study of cosmic dust (IDPs) and potential source materials. In Lunar and Planetary Science XVII, pp. 260-261. Lunar and Planetary Institute, Houston.

Kissel J., Sagdeev R.Z., Bertaux J.L., Angarov V.N., Audouze J., Blamont J.E., Büchler K., Evlanov E.N., Fechtig H., Fomenkova M.N., von Hoerner H., Inogamov N.A., Khromov V.N., Knabe W., Krueger F.R., Langevin Y., Leonas V.B., Levasseur-Regourd A.C., Managadze G.G., Podkolzin S.N., Shapiro V.D., Tabaldeyev S.R., and Zubkov B.V. (1986a) Composition of comet Halley dust particles from Vega observations, Nature, 321, 280-282.

Kissel J., Brownlee D.E., Büchler K., Clark B.C., Fechtig H., Grün E., Hornung K., Igenbergs E.B., Jessberger E.K., Krueger F.R., Kuczera H., McDonnell J.A.M., Morfill G.M., Rahe J., Schwehm G.H., Sekanina Z., Utterback N.G., Völk H.J., and Zook H.A. (1986b) Composition of comet Halley dust particles from Giotto observations, Nature, 321, 336-337.

Sommer M.A., and Gibson E.K. (1985) Determination of volatile components of fluid inclusions in Archean rocks using micro-crushing with a capacitance manometer and mass spectrometer. In Lunar and Planetary Science XVI, pp. 801-802. Lunar and Planetary Institute, Houston.

Sommer M.A., and Gibson E.K. (1986) Volatile determinations of individual fluid inclusions within the 3.4 B.Y. North Pole barites from the Warrawoona group, Northwestern Australia. In Lunar and Planetary Science XVII, pp. 815-816. Lunar and Planetary Institute, Houston.

Zolensky M.E., McKay D.S., Schramm L.S., VerPloeg K.L., Warren J.L., and Watts L.A. (1986) Cosmic Dust Catalog 8, 1. JSC 22425, NASA Johnson Space Center, Houston.

DIGITAL SIGNAL PROCESSING ALGORITHMS FOR AUTOMATIC VOICE RECOGNITION

Final Report

NASA/ASEE Summer Faculty Fellowship Program - 1987

Johnson Space Center

Prepared by:	Nazeih M. Botros, Ph.D.
Academic Rank:	Assistant Professor
University & Development:	Southern Illinois University Department of Electrical Engineering Carbondale, Illinois 62901
NASA / JSC	
Directorate:	Engineering
Division:	Tracking & Communication
Branch:	Telemetry & Audio
JSC Colleague:	Bill Jordan
Date:	August 7, 1987
Control Number:	NGT 44-001-800

ABSTRACT

The main objective of this research is to investigate the current digital signal analysis algorithms that are implemented in automatic voice recognition algorithms. Automatic voice recognition means, in simple terms, the capability of a computer (machine) to recognize and interact with verbal commands. In this research I focus on the digital signal, rather than the linguistic, analysis of speech signal. Several digital signal processing algorithms are available for voice recognition. Some of these algorithms are: Linear Predictive Coding (LPC), Short-time Fourier Analysis, and Cepstrum Analysis. Among these algorithms, the LPC is the most widely used. This algorithm has short execution time and do not require large memory storage. However, it has several limitations due to the assumptions used to develop it. The other two algorithms are frequency-domain algorithms with not many assumptions, but need longer execution time and larger storage, consequently they are not widely implemented or investigated. However, with the recent advances in the digital technology, namely the high density memory chips and the ultra fast digital signal processors, these two frequency-domain algorithms may be investigated in order to implement them in voice recognition. This research is concerned with real-time, microprocessor-based recognition algorithms.

INTRODUCTION

For more than a decade the United States Government, Foreign countries especially Japan, private corporations, and universities have been engaged in extensive research on human-machine interaction by voice. The benefits of this interaction is especially noteworthy in situations when the individual is engaged in such hands/eyes-busy task, or in low light or darkness, or when tactile contact is impractical/impossible. These benefits make voice control a very effective tool for space-related tasks. Some of the voice control applications that have been studied in NASA-JSC are: VCS Flight experiments, payload bay cameras, EVA heads up display, mission control center display units, and voice command robot. A special benefit of voice control is in zero gravity condition where voice is a very suitable tool in controlling space vehicle equipment.

Automatic speech recognition is carried out mostly by extracting *features from the speech signal and store them in reference templates in the computer. These features carry the signature of the speech signal. These reference templates contain the features of a phoneme, word, or a sentence, depending on the structure of the recognizer. If a

voice interaction with the computer takes place, the computer extracts features from this voice signal and compares it with the reference templates; if a match is found, the computer executes a programmable task such as moving the camera up or down.

A speaker-dependent recognizer is the one that is customized to a particular speaker. The template of this particular speaker is stored in the recognizer memory as the reference template; only this speaker can use that recognizer. A speaker-independent recognizer can be used by any speaker assuming that the speaker's language and dialog are the same of the recognizer. An isolated-word recognizer is the one that can recognize only a single word at a time; A pause should be inserted by the speaker between words. A connected-word recognizer is the one that can recognize a string of spoken words, no pauses needed. A recognizer can be built to recognize the word(s) of the speaker or can recognize the identity of the speaker from his spoken words.

At present time most of the commercially available recognizers are speaker-dependent,isolated word recognition, with limited vocabulary. "Current speech recognition technology is not sufficiently advanced to achieve high performance on continuous spoken input with large vocabularies and or arbitrary talkers." One of the factors that limits the performance of the current recognizers is the

efficiency of the recognition algorithms. "Significant research efforts are required in the design of algorithms and systems for the recognition of continuous speech in complex application domains, for speaker-independent operation, and for robust performance under conditions of degraded input."[1]

Several digital signal processing algorithms are available for speech feature extraction. The efficiency of the current algorithms is limited by: hardware restriction, execution time, and easiness of use. Some of these algorithms are: Linear Predictive Coding (LPC), Short-time Fourier Analysis, and Cepstrum analysis. Among these algorithms, the LPC is the most widely used since it is easy to use, short execution time, and do not require large memory storage. However, this algorithm has several limitations due to the assumptions that the algorithm is based upon. The other two algorithms usually need longer execution time and larger storage and consequently they are not widely implemented or investigated.

[1] Extraction from the National Research Council report, December 12, 1984.

FINDINGS

In this section we discuss four digital algorithms that been implemented in automatic voice recognition. These algorithms extract a certain number of feature from the speech signal. These features carry the signature of the signal. Recognition is achieved by comparing the feature set of the speaker with the reference feature set. A decision rule is implemented to decide on whether the speaker feature set matches the reference set or not.

1. Filter Bank-Analysis of Speech, [4].

In this algorithm the feature set consists of the speech energy within a certain number of frequency bands. The frequency range of the speech signal is divided into bands. This number varies from 5 to as many as 32, and the spacing between the bands is normally linear until about 1000 Hz, and logarithmic beyond 1000 Hz.

The energy within each frequency band is measured as shown in Figure 1. The speech signal passes through a bank of band pass filters; each filter covers a certain frequency band. The output of the pass band filter passes through nonlinear circuit, such as square law detector or full wave rectifier, and a low pass filter. The output of the nonlinear circuit is proportional to the square of the

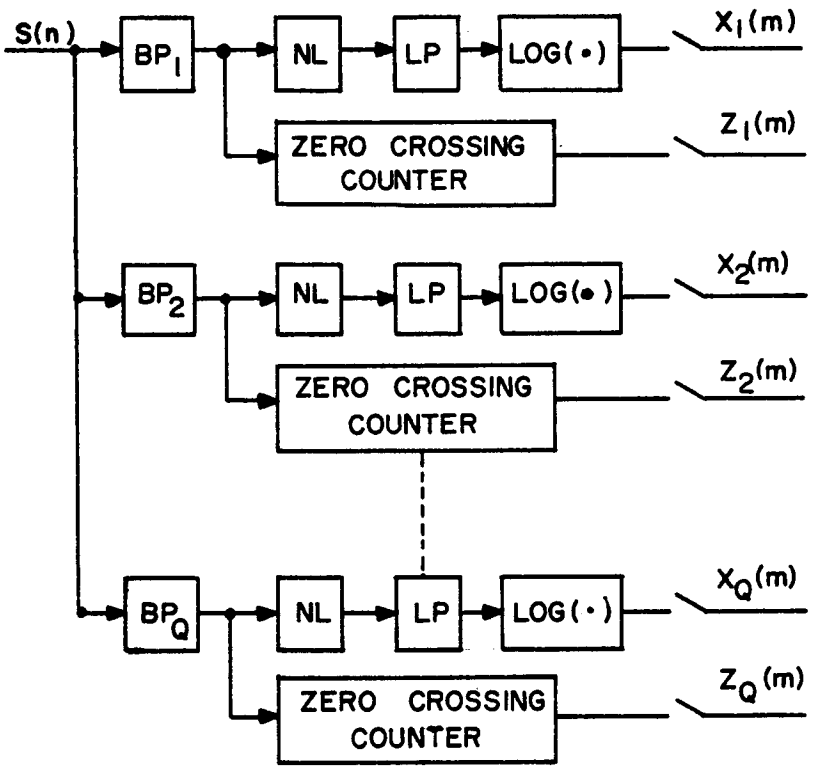


FIGURE 1. FILTER BANK - ALGORITHM.

amplitude of the signal and hence can be taken as a measure of the speech energy in this band. A logarithmic circuit is used to reduce the dynamic range of the intensity signal and the compressed output is digitized with a sample rate of twice the low pass filter cutoff frequency.

The low pass filter cutoff frequency is typically about 20-30 Hz; Accordingly the sample rate of the digitizer is selected to be from 40 to 60 Hz. If the number of the band pass filters is 5 and the sample rate is 40, then the number of features (the energy per frequency band) to represent a 1 second of the speech signal is 200. If we sampled the raw speech signal without using the filter bank, then the number of features will be 1000 for sample rate of 10 KHz. So by using filter bank-analysis the number of features is reduced by a rate of 50 to 1. For many recognizers, this feature set is supplemented by adding the number of times the signal crosses the zero time axis. This number of this zero crossing is related to the frequency pitch of the speech signal.

2. Linear Predictive Analysis, [1-7].

This algorithm is built on the fact that there is a high correlation between adjacent samples of the speech in the time domain. This fact means that an nth sample of speech signal can be predicted from previous samples. The

correlation can be put in a linear relationship and we get what is called Linear Prediction Model. This relationship can be written as:

$$\hat{Y}_n = a_1 Y_{n-1} + a_2 Y_{n-2} + \ldots + a_p Y_{n-p} \qquad \ldots\ldots\ldots\ldots (1)$$

where p is the order of analysis; usually p ranges from 8 to 12. y_n is the predicted value of speech at time n and a's are the linear predictive coefficients. The error E_n that resulted from the above linear relationship is:

$$E_n = Y_n - \hat{Y}_n = Y_n - \left(\sum_{i=1}^{p} a_i Y_{n-i} \right) \qquad \ldots\ldots\ldots\ldots (2)$$

The error E_n is called the prediction error. Setting $-a_i$ as a_i, the prediction error becomes:

$$E_n = Y_n + \sum_{i=1}^{p} a_i Y_{n-i} \qquad \ldots\ldots\ldots\ldots (3)$$

$$= \sum_{i=0}^{p} a_i Y_{n-i}, \qquad a_o = 1. \qquad \ldots\ldots\ldots\ldots (4)$$

Squaring the above Equation and taking the average:

$$\overline{E_n^2} = \overline{(Y_n + a_1 Y_{n-1} + a_2 Y_{n-2} + \ldots + a_p Y_{n-p})^2} \qquad \ldots\ldots\ldots\ldots (5)$$

To find the predictive coefficients which give minimum predictive error, the above Equation is partially

differentiated with respect to a's and time average term by term is taken:

$$\begin{bmatrix} r_0 & r_1 & r_2 & \cdots, & r_{p-1} \\ r_1 & r_0 & r_1 & \cdots, & r_{p-2} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ r_{p-1} & r_{p-2} & r_{p-3} & \cdots, & r_0 \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_p \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ \cdot \\ \cdot \\ \cdot \\ r_p \end{bmatrix}, \quad \text{.....(6)}$$

with

$$r_0 = \overline{Y_n Y_n}, \qquad r_j = \overline{Y_n Y_{n+j}} = \overline{Y_{n-j} Y_n} \quad \text{.....(7)}$$

where r_j is a correlation coefficient of waveform $\{y_n\}$ and $r_{-j} = r_j$ by the assumptions of stationary state of y_n The coefficients a_i's exist only if the matrix in Equation 6 is positive definite. To ensure that this condition is satisfied, y_n is multiplexed by a time window W_n. This multiplexing makes y_n exists in a finite interval from 0 to N-1, where N is the interval of the Window; a stable solution for Equation 6 is always obtained. Accordingly, r_j is written as:

$$r_j = \frac{1}{N} \sum_{n=0}^{N-j-1} Y'_n Y'_{n+j} \quad \text{.....(8)}$$

$$= \frac{1}{N} \sum_{n=0}^{N-j-1} Y_n Y_{n+j} W_n W_{n+j} \quad \text{.....(9)}$$

Calculation of the correlation coefficients by window multiplexing is called the correlation method. Some recognizers use the correlation coefficients as the feature set. However for full LPC analysis a_i's are calculated by solving Equation 6.

The LPC model represents an all-pole model. The relation between the input x_i and the output y_n of this system is written as:

$$Y_n + \sum_{i=1}^{p} a_i Y_{n-1} = X_n, \quad \text{.....(10)}$$

Equation 10 is called the auto-regressive process. The system function H(z) can be written as:

$$H(z) = 1/(1+a_1 z^{-1}+.....+a_p z^{-p}) \quad \text{.....(11)}$$

a_i's correspond to the resonance frequencies of the signal and if p, the order of the analysis is selected correctly, these a_i's represent the formants, frequencies at which peaks of the power spectrum of the speech signal occur.

A block diagram representing an algorithm for voice recognition based on LPC analysis is shown in Figure 2. As shown in this Figure, the digitized data is divided into frames each of length N. The distance between consecutive frames is M. If N=M then there is no overlapping between

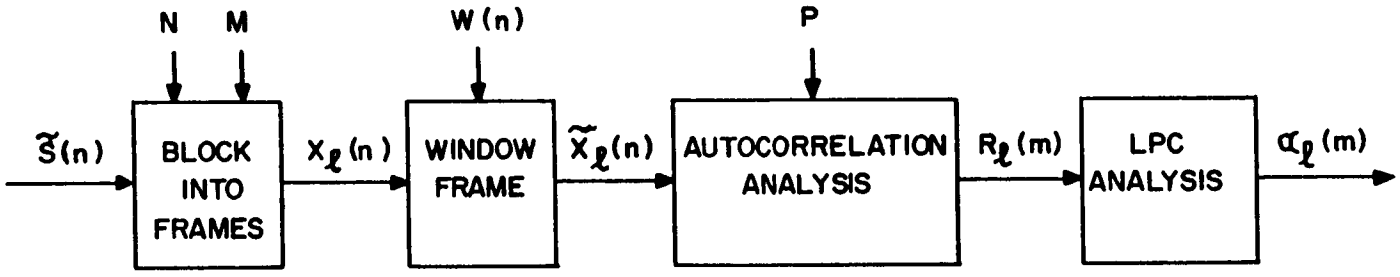


FIGURE 2. LPC-BASED ALGORITHM FOR AUTOMATIC VOICE RECOGNITION

frames; if M<N, then there is overlapping between frames. Typical values of N are from 100 to 500 data points. For large number of N the analysis is called wideband analysis and for small values of N the analysis is called narrow band analysis. Because the rate of speaking of any subject is not fixed, it changes with time, a time wrapping algorithm is used to take this fact into consideration.

Limitations of the LPC analysis

The LPC analysis needs relatively small memory storage and has a short execution time. On the other hand to apply this analysis to any system, the system should satisfy the following conditions:

a. The system is an all-pole system. Speech system does not explicitly satisfies this condition. However, any system can be approximated to all-pole system by increasing the number of poles relative to the zeros in the system.

b. The input to the system is either a single impulse or pure white noise. This is not explicitly true especially in the case of a female voiced sound where the pitch period is generally short.

c. The system is time-invariant. The speech system is time varying system, however using a window can approximate the system to a time-invariant system.

3. Short-Time Fourier Analysis, [2].

In this analysis the domain of the signal is transformed from time domain to frequency domain. Frequency domain analysis is more desirable than the time domain for the following reasons:

a. In the frequency domain the signal is decomposed into its frequency components. Investigation of these frequency components lead to understanding the nature of the signal as well as the effect of the noise on it.

b. The input/output relation of any system in the frequency domain is the product of the Fourier transform of the impulse response of the system and the Fourier transform of the input. In the time domain this relationship is a convolution which is more complicated than multiplication.

c. The autocorrelation function of the system, which is often used to describe the statistical properties of the signal is related in a simple relationship with the power spectrum of the signal.

Since the speech signal is a time-varying signal; the spectrum of the signal changes with time, then Fast Fourier Transform can not be applied directly. Short time Fourier transform should be applied instead of FFT. The short time Fourier transform of a time domain signal, x(m), can be written as:

$$X_n(e^{jw}) = \sum_{m=-\infty}^{\infty} W(n-m)x(m)e^{-jwm} \qquad \text{.....(12)}$$

where W(n-m) is a time window. The window is used here to justify the use of Discrete Fourier transform for a time-varying signal. The Hamming window is the most widely used; it can be written as:

$$W(n) = 0.54 - .46\cos(2\pi n/(N-1)); \quad 0 \leq n \leq N-1 \qquad \text{.....(13)}$$
$$= 0 \quad ; \quad \text{elsewhere}$$

As is shown in Equation 12, short time Fourier includes both convolution process and DFT process. Calculation of short time Fourier by Equation 12 takes a very long execution time. It can be shown that Equation 12 can be written as:

$$X_n(e^{jw}) = e^{-j\frac{2\pi}{N}kn} \sum_{q=0}^{N-1} \{ \sum_{r=-\infty}^{\infty} x_n(Nr+q)\} e^{-j\frac{2\pi}{N}kq} \qquad \text{.....(14)}$$

$$X_n(e^{jw}) = e^{-j\frac{2\pi}{N}kn} \sum_{q=0}^{N-1} U_n(q)e^{-j\frac{2\pi}{N}kq} \qquad \text{.....(15)}$$

where

$$U_n(q) = \sum_{r=-\infty}^{\infty} X_n(Nr+q) \qquad \text{.....(16)}$$

As is shown in Equation 15, the short time Fourier Transform can be written as an FFT process multiplied by : $e^{-j\frac{2\pi}{N}kn}$;the execution time is relatively short and consequently Equation 15 is used to calculate the short time

Fourier Transform. After calculation of the short time Fourier Transform, the spectrum can be investigated to extract features, such as formants, that can represent the speech signal.

4. Cepstrum Analysis, [1],[2].

The speech waveform can be approximated by:

$$s(t) = \sum_{n=-\infty}^{\infty} s_0(t-nT) = s_0(t) * [\sum_{n=-\infty}^{\infty} \delta(t-nT)] \quad \ldots\ldots\ldots(17)$$

where $s_0(t)$ is impulse response of the speech generating system and $\sum_{n=-\infty}^{\infty} \delta(t-nT)$ is the pulse train with period T. Applying Fourier Transform to both sides of the above Equation, then:

$$S(w) = S_0(w)\{\sin[(2N+1)\tfrac{1}{2}wT]/\sin\tfrac{1}{2}wT\}^2 \quad \ldots\ldots\ldots(18)$$

Where S(w) and $S_0(w)$ are the power spectra of s(t) and $s_0(t)$ respectively. Taking the logarithm of both sides of the above Equation, then:

$$\log_e S(w) = \log_e S_0(w) + 2\log_e\{(2N+1)\tfrac{1}{2}wT]/\sin\tfrac{1}{2}wT\} \quad \ldots\ldots\ldots(19)$$

The first term on the right-hand side of the above Equation represents a relatively slow change in frequency (the speech generating system) and the second term represents

a relatively high change in frequency with fundamental frequency $2\pi/T$. This means that the above Equation consists of two separable terms with respect to frequency. By taking the Inverse Fourier Transform, we can have two terms; the first term corresponds to the spectrum envelope and the second term corresponds to the pitch excitation. The result of this inverse Fourier transformation is called cepstrum and the variable corresponding to frequency is called quefrency. Figure 3 demonstrates the cepstrum analysis.

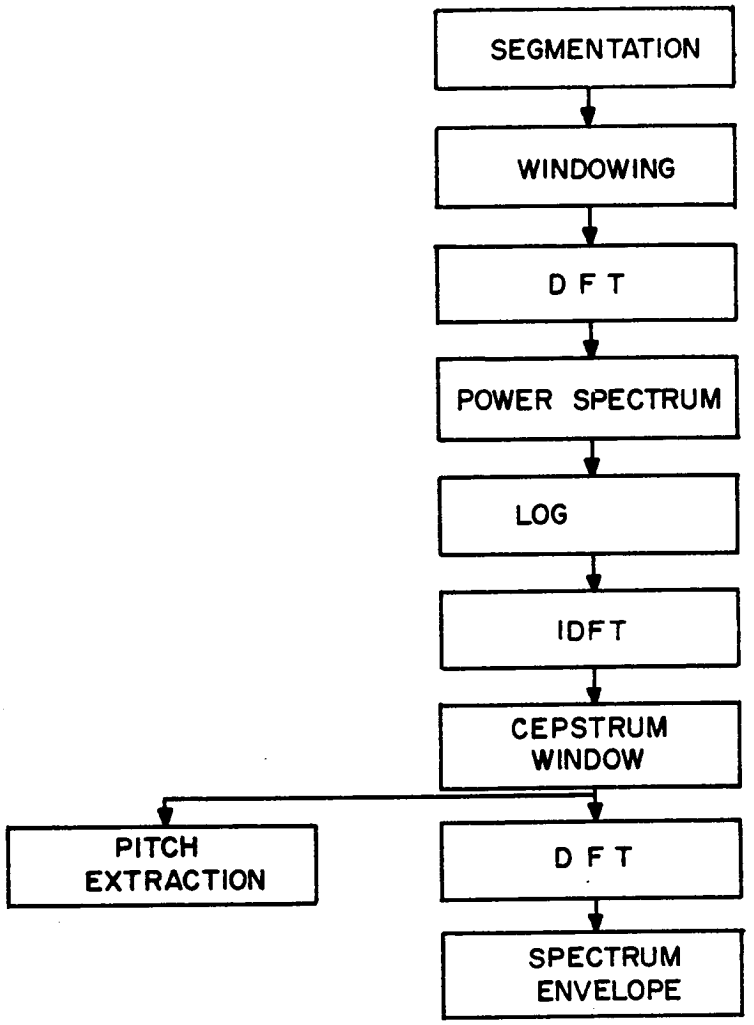


FIGURE 3. CEPSTRUM ANALYSIS OF SPEECH

CONCLUSION

Four digital signal processing algorithms for automatic speech recognition have been discussed in this research. The Linear Predictive Analysis is the most widely used for its low memory storage requirement and its short execution time. However this algorithm has several limitations. The Short Time Fourier Analysis and the Cepstrum Analysis are frequency-domain algorithms; these two algorithms require large memory storage and their execution time is relatively long. These two frequency domain algorithms do not have as many limitations as of the LPC analysis. The best approach to introduce a high performance algorithm for automatic voice recognition is a combination of LPC analysis and the frequency domain Analysis.

REFERENCES

1. Saito, S. and Nakato, K., 1985, Fundamental of Speech Signal Processing, Shuzo Saito and Kazuo Nakato, Academic Press, Inc.
2. Fallside,F. and Woods, W., 1985, Computer Speech Processing, Prentice Hall International (U.K.) Ltd.
3. Itakura, F., 1975, "Minimum Prediction Residual Principle Applied to Speech Recognition", IEEE Trans. Acoust., Speech, Signal, Processing, Vol. ASSP-23, pp. 67-72.
4. Rabiner, L. R.,1978, "On Creating Reference Template for Speaker Independent Recognition of Isolated Word", IEEE Trans. Acoust., Speech, Signal, Processing, Vol. ASSP-26, pp. 34- 42.
5. Rabiner, L. R. and Levinson, S. E., 1981, "Isolated and Connected Word Recognition-Theory and Selected Application", IEEE Trans. on Communication, Vol. comm. 29, No. 5.
6. Rabiner, L. R., Levinson, S. E., Rosenberg, A. E., and Wilpon, J. G., 1979, "Speaker-Independent Recognition of Isolated Word Using Clustering Technique", IEEE Trans. Acoust., Speech, Signal, Processing, Vol. ASSP-27, pp. 336-349.

Shuttle Computational Grid Generation

Final Report

NASA/ASEE Summer Faculty Fellowship Program - 1987

Johnson Space Center

Prepared by:	Ing Chang, Ph.D.
Academic Rank:	Associate Professor
University and Department:	Prairie View, Texas 77446 Department of Mechanical Engineering Prairie View, Texas 77446
NASA/JSC	
Directorate:	Engineering
Division:	Advanced Programs Office
Branch:	Aeroscience
JSC Colleague:	Chien-Peng Li, Ph.D.
Date:	August 21, 1987

ABSTRACT

The well known Karman-Trefftz conformal transformation, consisting of repeated applications of the same basic formula, were found to be quite successful to body, wing, and simple wing-body cross sections in recent years. In this report, it is intended to extend this grid generation technique to cross sections of more complex forms, and also intended to make the grid generation process more automated.

Computer programs were written for the selection of "hinge points" on cross section with angular shapes, the Karman-Trefftz transformation of arbitrary shapes, and the special transform of hinge point on the imaginery axis.

The present work is served as a feasibility study for the future application of conformal mapping grid generation to complex three dimensional configurations. Examples such as Orbiter vehicle section and a few others were used in the present study. Computer programs are not included in this report.

INTRODUCTION

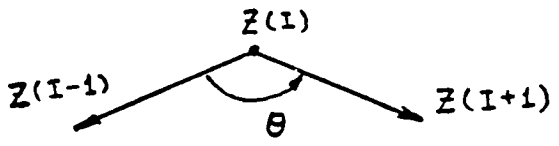
Conformal mapping is one of the several techniques used in numerical grid generation (1-2). For shapes with corners, the Karman-Trefftz transformation is commonly used to smoothen the angle while mapping the physical configuration into near circle on the computational plane. Grids can be easily generated on computational plane, and inverse transform will map the grids back to the physical plane. On these grids, the appropriate numerical methods can be applied to solve the flow dynamics problems. Just a few are listed in the reference (3-4). Care must be maintained for using those transforms and for complex geometries, difficulties could be encountered.

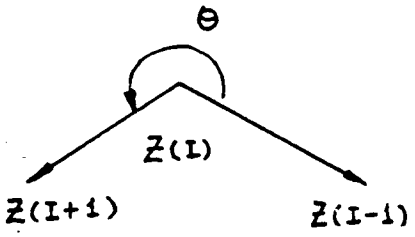
In the present work, it is intended to extend and automate the transforms for any arbitrary shapes. This should include the hinge points selections, the Karman-Trefftz with arbitrary branch cuts, automatic grid generation on computational plane, and the inverse transforms to the physical plane. Computer programs were written but not included in this report, the results are shown graphically. For these graphics, the vertical axis and horizontal axis are not at exactly the same scale, so, there is a slight distortion for the figures. To show the sequence of transformation, the figures on the same page should be read from left to right and from top to bottom.

LOCATE THE HINGE POINTS

Hinge points are used in the Karman-Trefftz transformation to smoothen the corners of the configurations to be transformed. They're located at these corners or very close to them. If a large number of coordinates are needed to describe the geometries, direct observations to locate the hinge points may be time consuming and less accurate. So, a simple program was written to automate this process of selection.

Refer to figure 1, for an arbitrary shape, a number of complex coordinates (Z (I), I = 1, N) are used to define the shape in the complex plane. Start from one end, Z (1), compute the angles at all the coordinates in sequence, compare with the pre-determined values of maximum deviations from 180°, the location of the hinge points (as well as the δ used in the transformation) can be selected by the computer. In our scheme, not all the locations exceed the limits are considered as hinge points, only the locations that represent the largest curvature locally are selected as hinge points. The computer program is proved to be successful on Shuttle sections as well as a few other geometries.





θ is computed from the following complex expression

$$\frac{Z(I+1) - Z(I)}{Z(I-1) - Z(I)}$$

If the angle computed is negative, add 2π to obtain θ

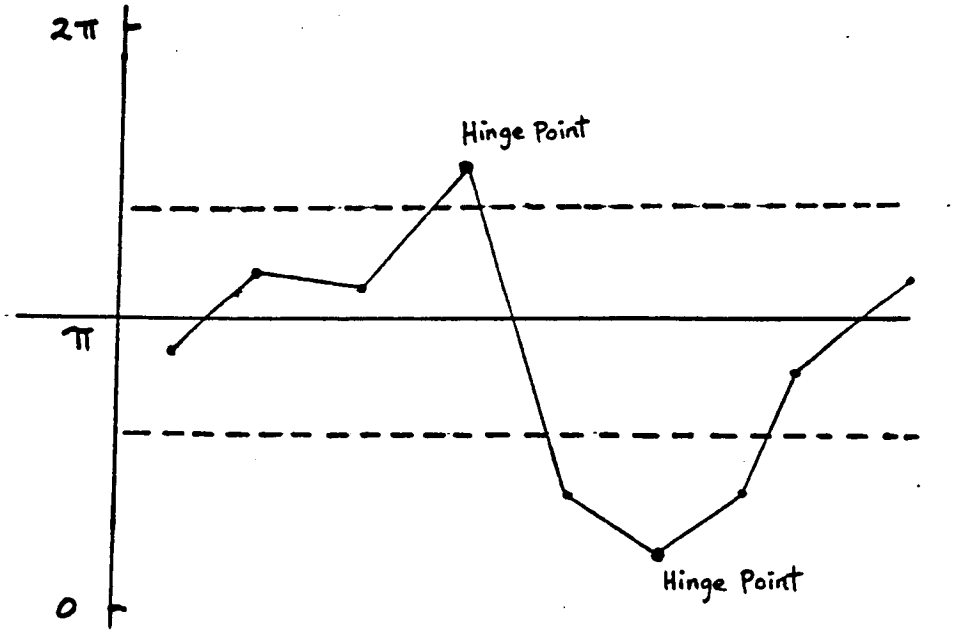


Fig. 1 Selection of hinge points

THE KARMAN-TREFFTZ TRANSFORMATION

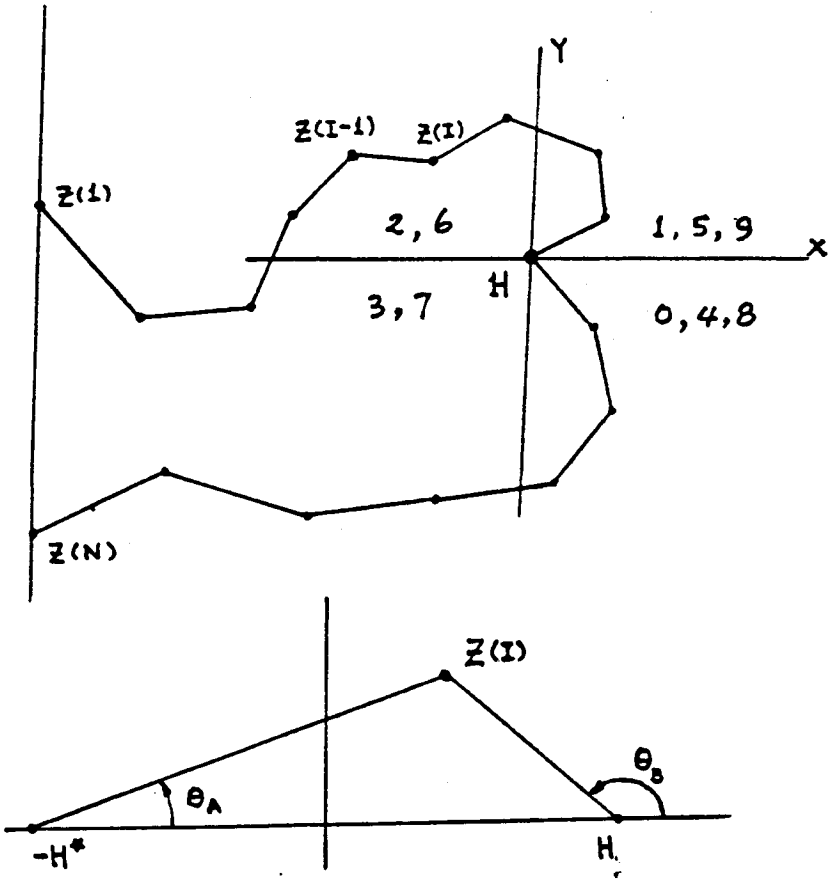
The Karman-Trefftz transformation is well known in the publications (5). In the present work, it is intended to write a program that can be applied to any hinge point not on the imaginery axis. For each hinge point, formulate the following Karman-Trefftz transforms

$$\frac{W(I)-1}{W(I)+1} = \left(\frac{Z(I)-H}{Z(I)+H} \right)^{\delta} \qquad I = 1, N$$

$$\delta = \frac{\pi}{2\pi-\alpha}$$

where
- $Z(I)$ = coordinates used to define the body
- $W(I)$ = transformed coordinates of the body
- α = measure of interior angle at hinge point
- H = coordinate of the hinge point
- H^* = complex conjugate of H

Refer to figure 2, at the hinge point, the plane is divided into four quarters but labeled them quarter 0 to quarter 9. Each coordinate of the body is in one particular quarter. The quarter numbers for coordinate Z (1) and
Z (N) can be obtained easily. The quarter number for other coordinates can be obtained by comparing with the adjacent coordinates. This should be done in sequence from Z (1) to Z (N) or from Z (N) to Z (1). It can also be done from
Z (1) to hinge point and then from Z (N) to hinge point. According to the general formulas mentioned in the figure, the Karman-Trefftz transformation can be completed. The difficult job of locating the branch cut is easily solved. Based on this concept, a computer program was written for the transformation. It proves to be successful as can be seen from a test example shown in figure 3. Other examples using this program are shown in the later sections.



1. IF sign of X_I = Sign of X_{I-1}
   and sign of Y_I = Sign of Y_{I-1} $\quad ID(I)=ID(I-1)$
2. IF sign of X_I = Sign of X_{I-1}
   and sign of $Y_I \neq$ sign of Y_{I-1} $\quad ID(I)=ID(I-1)+(-1)^{ID(I-1)}$
3. IF sign of $X_I \neq$ sign of X_{I-1}
   and sign of Y_I = Sign of Y_{I-1} $\quad ID(I)=ID(I-1)-(-1)^{ID(I-1)}$

$$\theta_B = (-2 + \frac{ID(I)}{2})\pi + ATAN(Y/X)$$

Fig. 2 Karman-Trefftz Transform

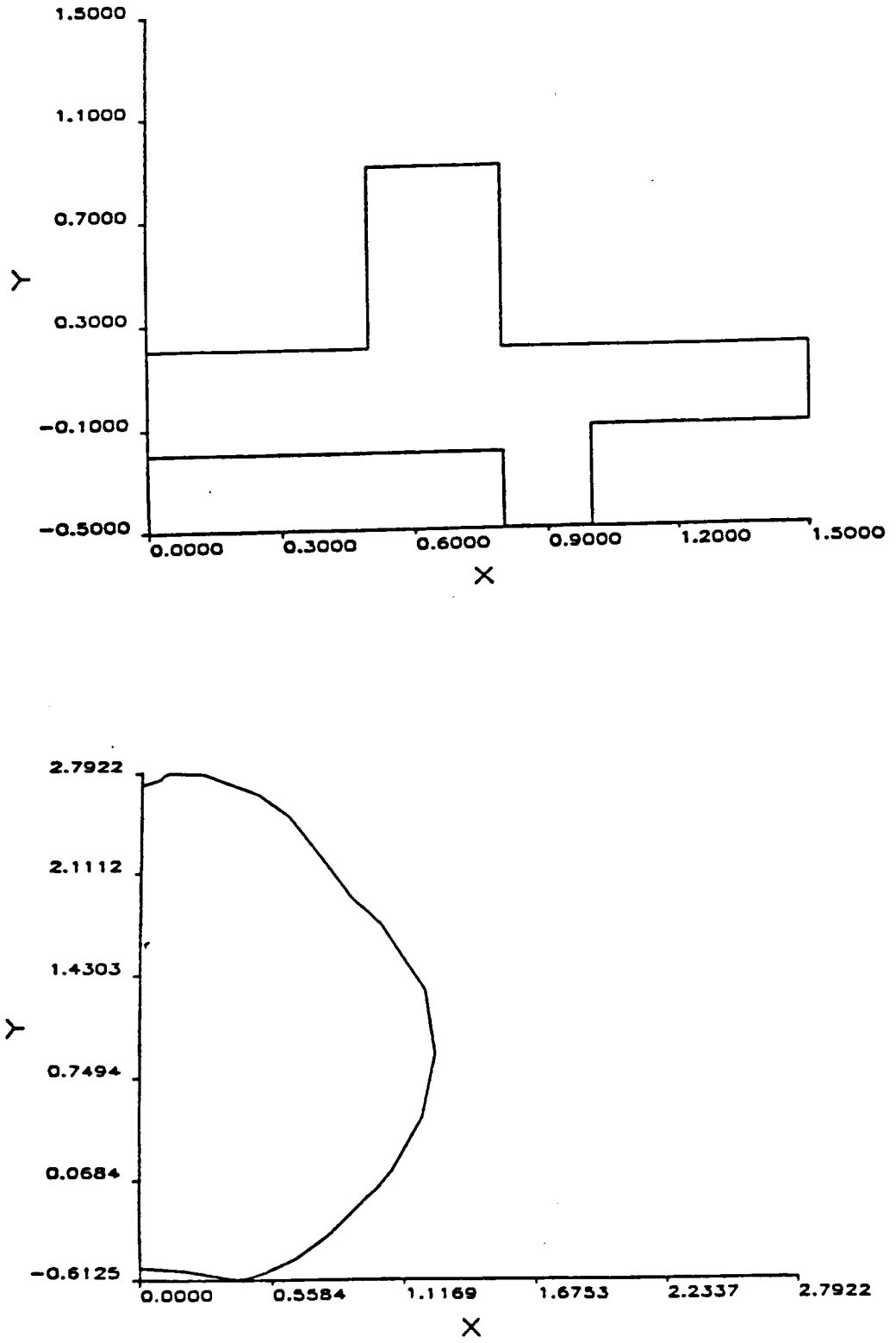


Fig. 3 Test example of Karman-Trefftz Transform from physical plane to computational plane

CONFORMAL MAPPING FOR HINGE POINT ON IMAGINERY AXIS

For hinge point on imaginery axis, the standard Karman-Trefftz transformation fails. But a modified transform formula can be used

$$\frac{W(I)-i}{W(I)+i} = \left(\frac{Z(I)-H}{Z(I)+H}\right)^{\delta}$$

$$\delta = \frac{\pi}{2\alpha}$$

α is shown in figure 4. This special transform should be applied to coordinate Z (1), if applied to coordinate Z (N), i in the formula should be replaced by -i. This transform can convert the angle at Z (1) or Z (N) to 90°. At the same time, if it is properly applied, it could lift the other hinge point on the imaginery axis. Special attention should be paid to the location of real axis (refer to figure 4), but within the allowable range, it has minor effect to the transformation as shown in figure 5. In figure 6 and figure 7, an example of rocket section is transformed in sequence from physical plane to computational plane. In this case, both the special transform and the standard Karman-Trefftz transforms are used. In figure 8 and figure 9, similar transformations were applied to an Orbiter vehicle section. Both transforms were done by computer programs. From these examples, it shows that with the suitable combination of both transforms, many problems with complex geometries can be solved.

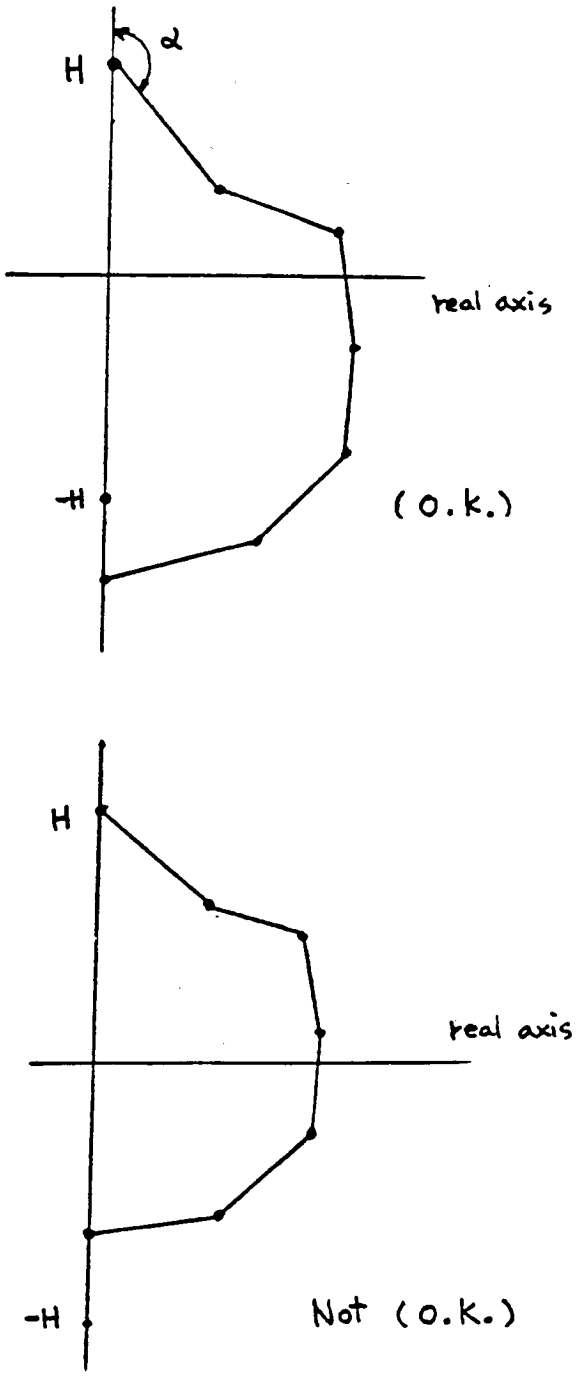


Fig. 4 Conformal mapping for hinge point on imaginery axis

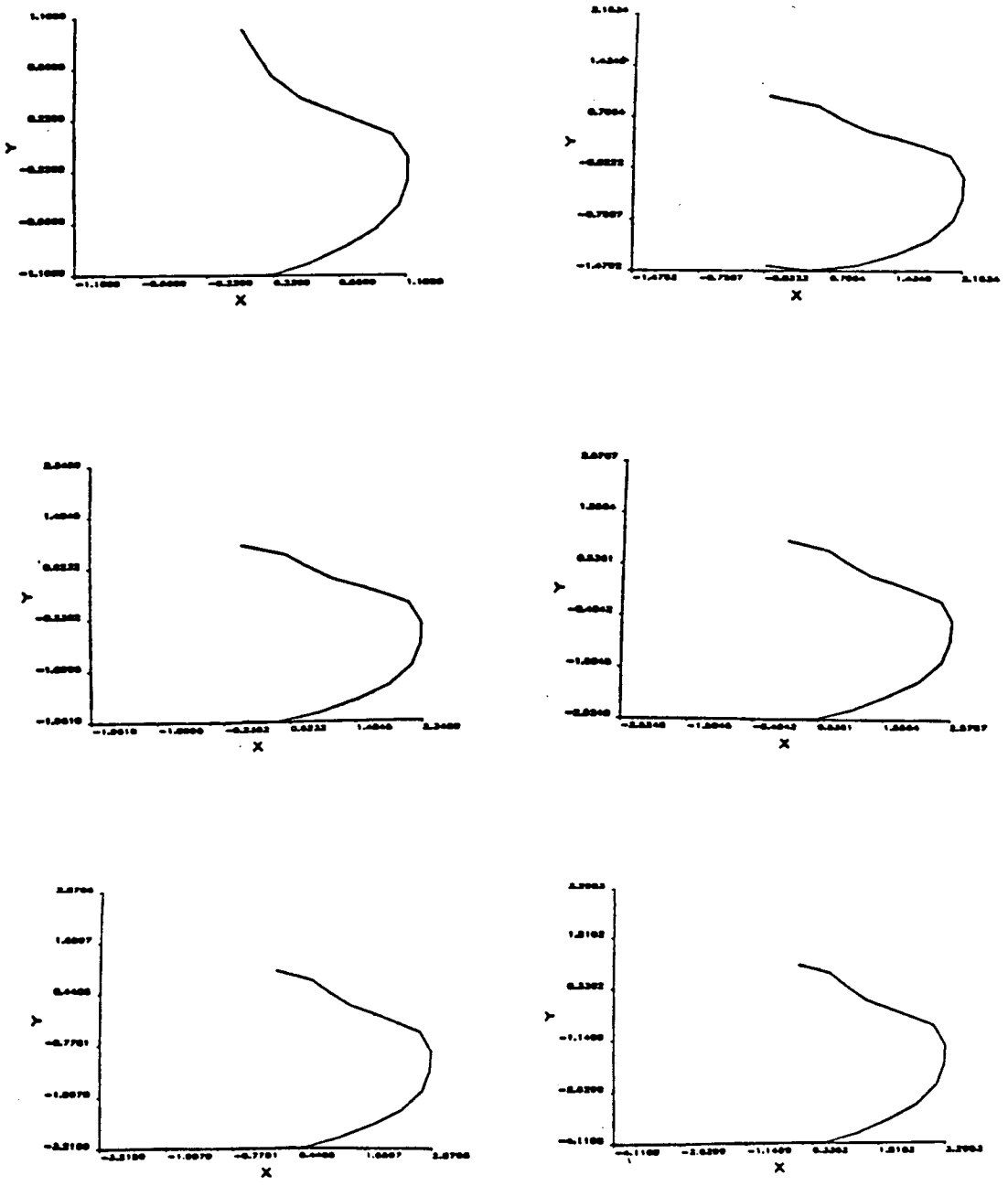


Fig. 5 Effect of real axis position variations
(First figure is before transformation.)

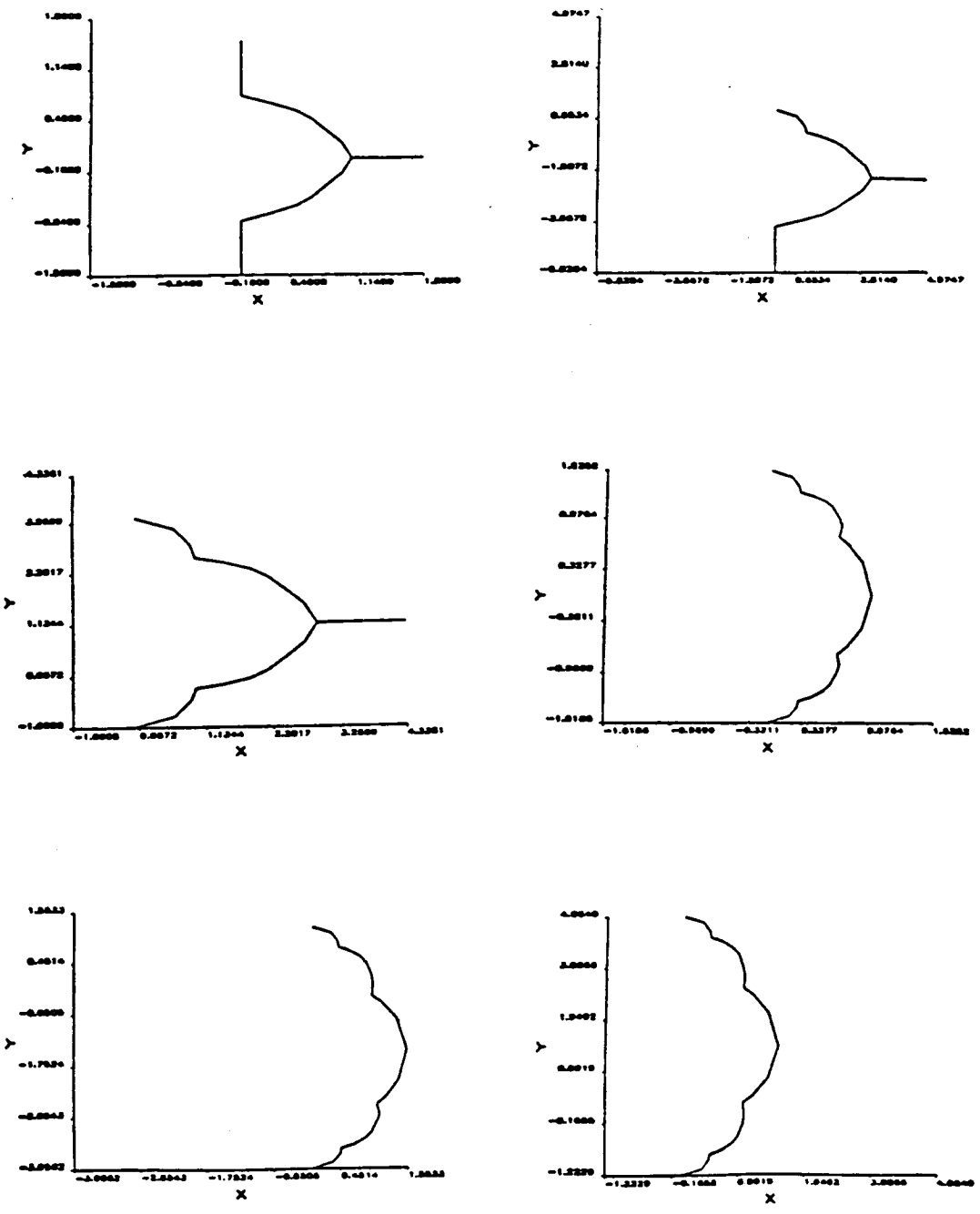


Fig. 6 The sequence to transform a finned rocket section into near circle in computational plane (continue in Fig. 7)

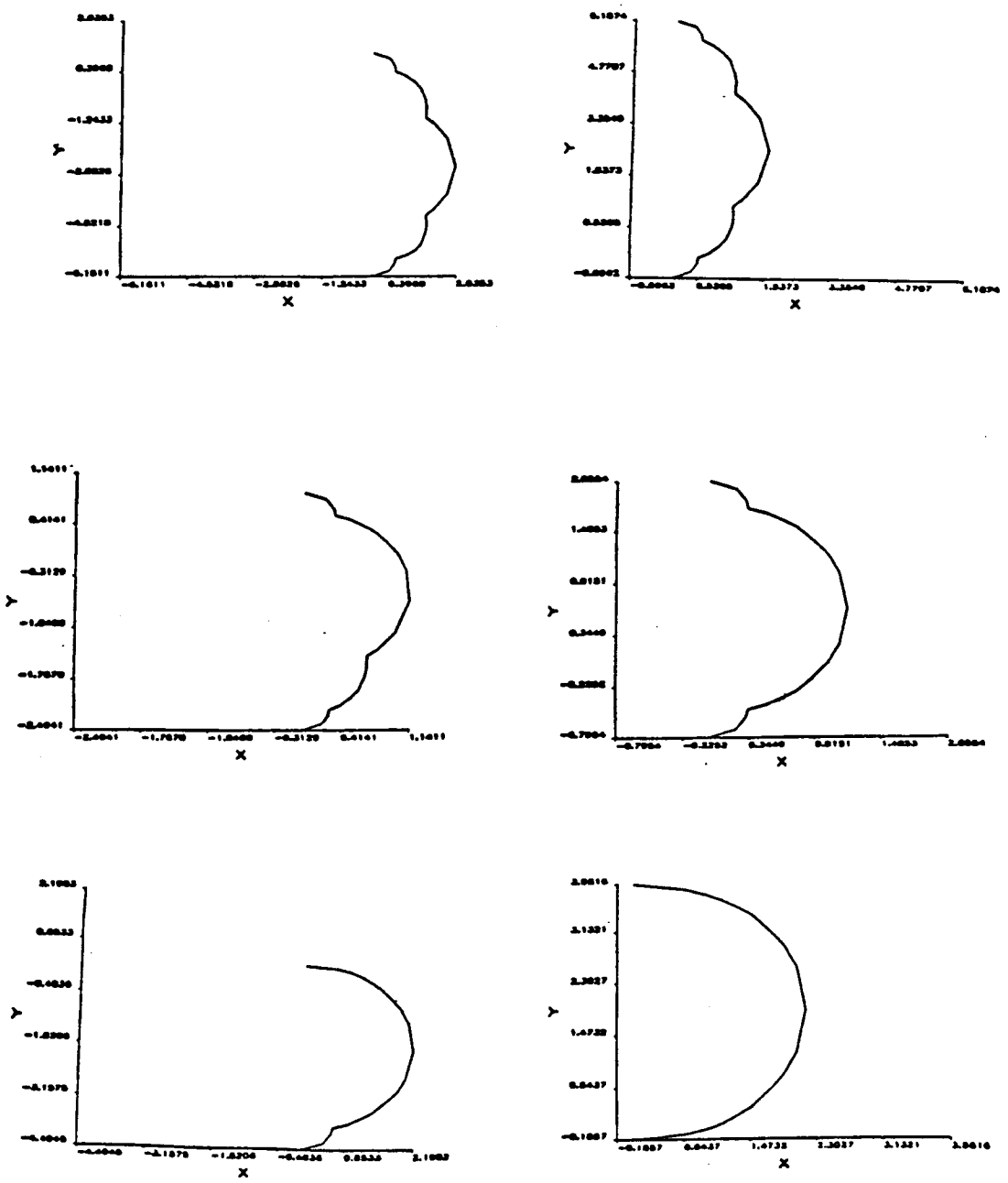


Fig. 7 Continuation of Fig. 6

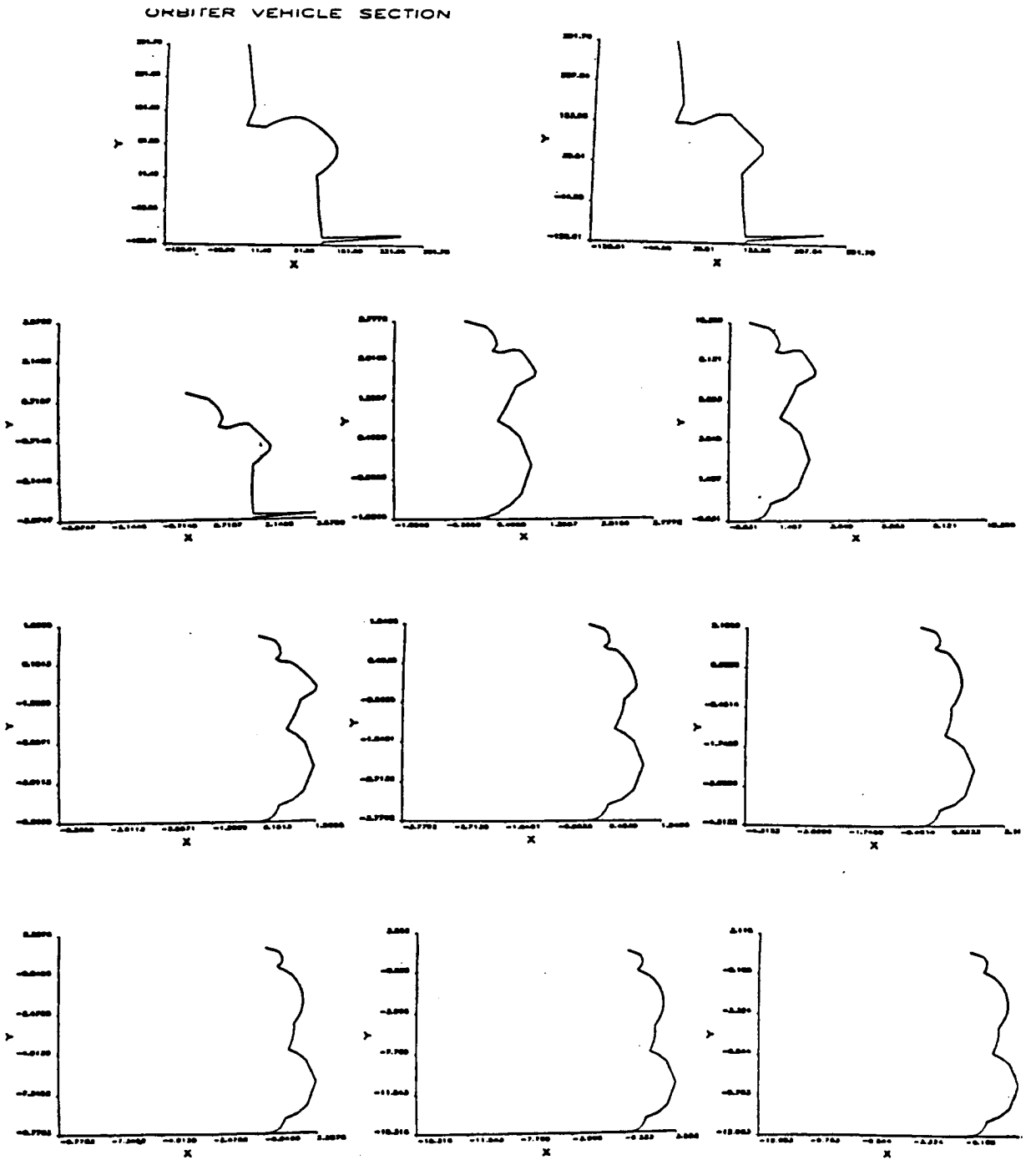


Fig. 8 Orbiter vehicle section transformation from physical plane to computational plane (continue in Fig. 9)

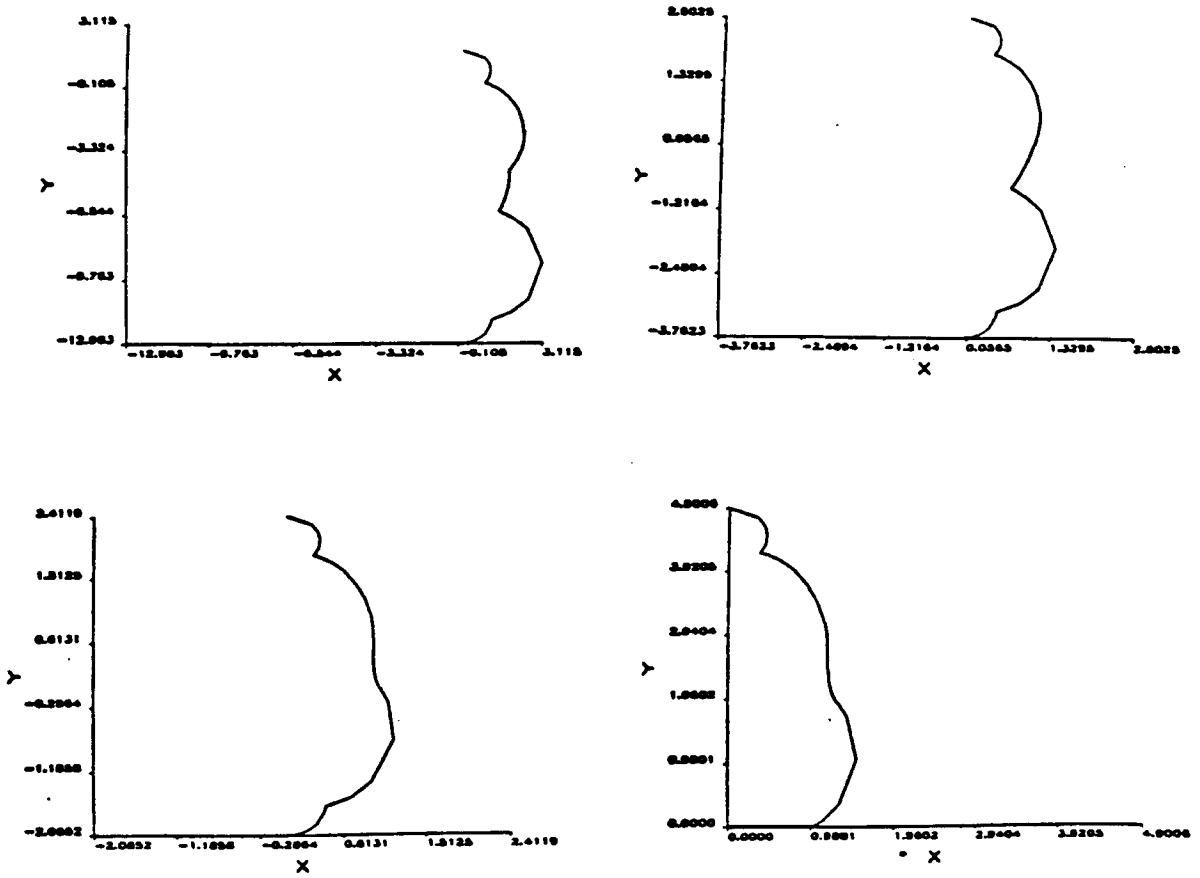


Fig. 9 Continuation of Fig. 8

GRID GENERATION ON COMPUTATIONAL PLANE AND THE INVERSE TRANSFORMATION

A simple program has been written to generate grids on computational plane. The program is written in such a way that there is always a line passing the hinge point. With minor changes, the programs used earlier can be used for inverse transformations. The grids on the computational plane can be mapped to the physical plane. In figure 10, the grids on an Orbiter vehicle section are shown on computational plane as well as on physical plane. In figure 11, that is for rocket section.

In conclusion, the computer programs developed in this study are quite general and can be used to solve complex shapes. But further study is needed to locate or generate hinge points automatically on a smooth but large curvature configuration.

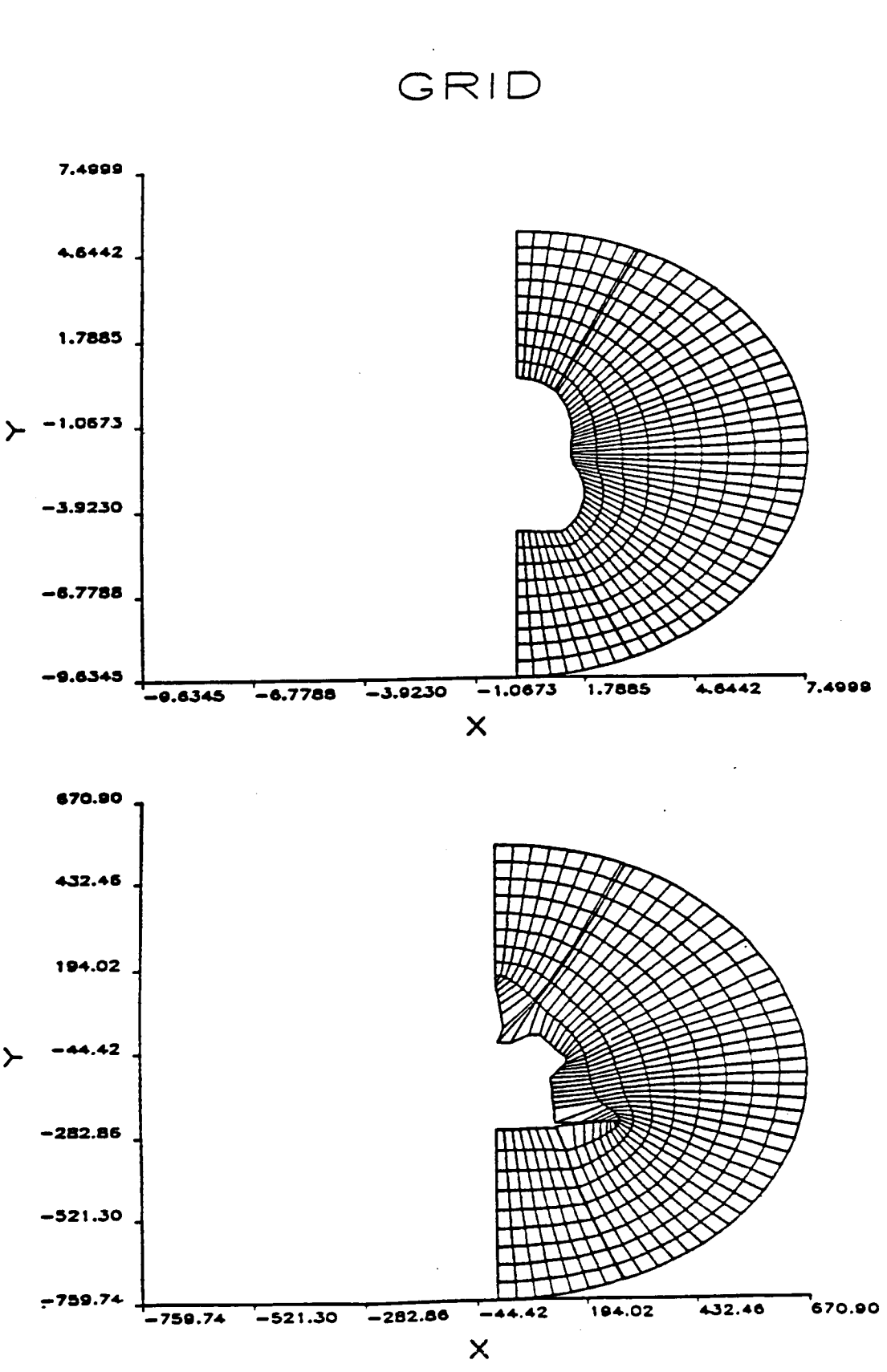


Fig. 10 Grids on an Orbiter vehicle section on computational plane and on physical plane

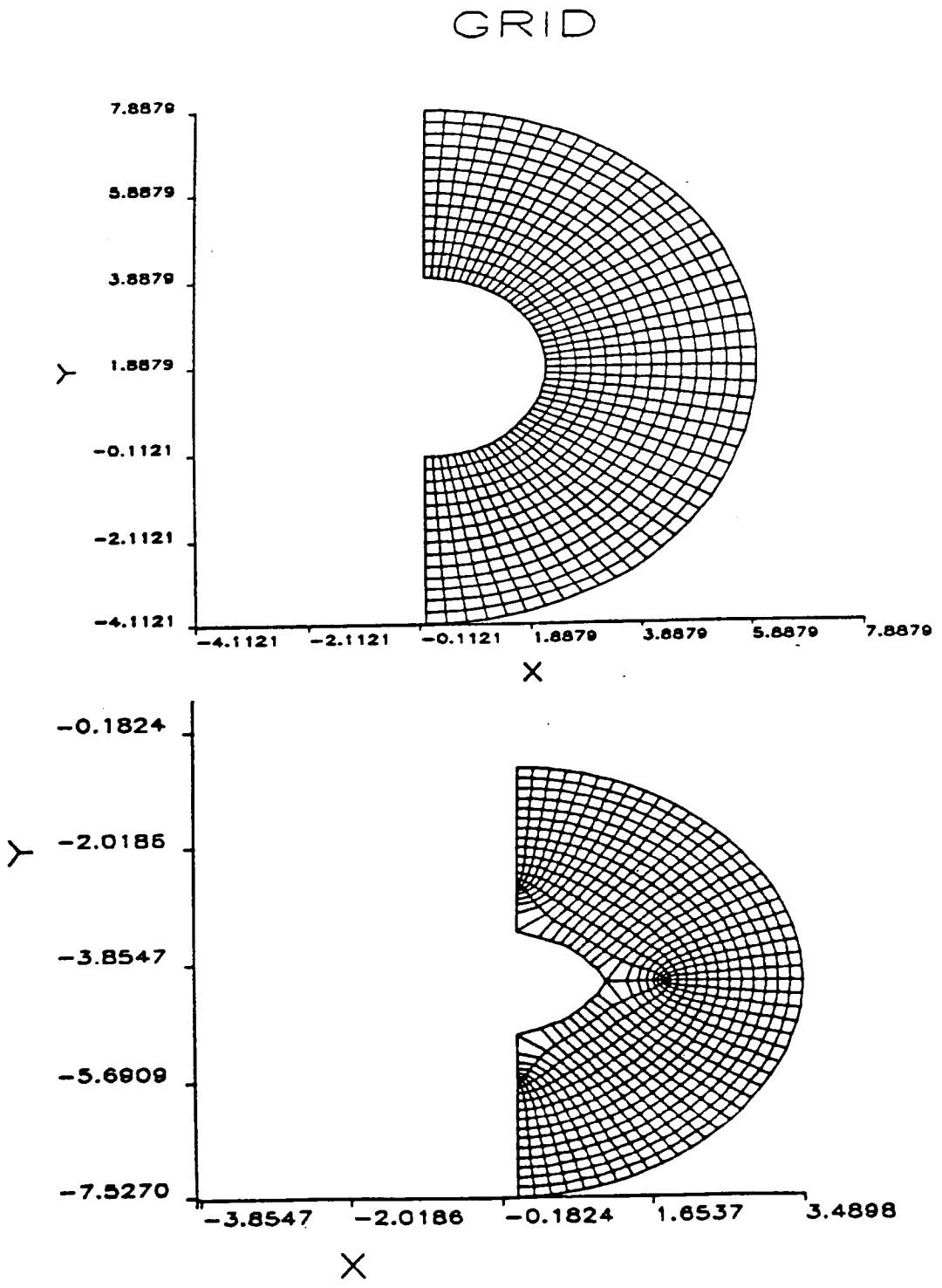


Fig. 11 Grids on Rocket Section

REFERENCES

1. Hall, D. W., "A Three-Dimensional Body-Fitted Coordinate System for Fluid Field Calculations on A Symmetric Nosetips," NASA Conf. Publ. 2166, 1980.

2. Anderson, D., et al, "Computational Fluid Mechanics and Heat Transfer," McGraw Hill, 1983.

3. Li, C. P., "Computation of Three-Dimensional Flow About Aerobrake Configurations," AIAA Aerospace Science Meeting, 1986.

4. Siclari, et al, "Application of NCOREL to Conical Multi-Finned and Multi-Faceted Configurations," AIAA Aerospace Science Meeting, 1986.

5. Moretti, G., "Conformal Mappings for Computations of Steady, Three-Dimensional, Supersonic Flows," ASME Publications, 13, 1976.

DEVELOPMENT OF A GRAPHICAL DISPLAY ON THE DMS TEST BED

FINAL REPORT

NASA/ASEE Summer Faculty Fellowship Program -- 1987

Johnson Space Center

Prepared by:	Robert A. Donnelly, Ph. D
Academic Rank:	Associate Professor
University & Department:	Auburn University Department of Chemistry Auburn University, AL 36849
NASA/JSC	
Directorate:	Engineering and Development
Division:	Avionics Systems
Branch:	Flight Data Systems
JSC Colleague:	Michael M. Thomas
Date:	August 26, 1987
Contract Number:	NGT 44-001-800

ABSTRACT

The DMS test bed is a model of a data network aboard Space Station. Users of the network share data relevant to the functional status of various systems (Guidance, Navigation and Control, Environmental Control and Life Support, etc.) aboard the Station. Users may inquire the status of myriad sensors, obtaining readings of Station subsystem status in real-time via the Data Acquisition and Distribution Service.

This project is aimed at development of a graphical display of the status of a simulation of the Environmental Control and Life Support System. Two broad issues were addressed: (1) Flexible, extensible software design; (2) The impact of utilizing standard processors, languages, and graphics packages implementing the software design concept.

Our experience gained with DEC hardware, the DEC implementation of the GKS graphics standard, and with Ada can be summarized as follows. DEC hardware seems adequate to the display tasks which form one part of the present research project. Highly graphics-intensive applications run slowly on the MicroVax GPX when it is programmed using the various DEC implementations of the GKS standard. Use of the GKS standard for graphics can provide code portability provided the FORTRAN/77 to GKS binding is used. Since no binding to Ada is currently available it makes sense to provide a procedural interface between Vax Ada and the FORTRAN/77 binding. This was accomplished, and can be used by other generations of graphics programmers who must program in the development environment provided by DEC.

I. INTRODUCTION

The DMS test bed is a model of a distributed communications network which is proposed for the Space Station. Users of the network have access to data generated by several subsystems likewise attached to the network. Users may inquire the status of these subsystems (and their components) through DADS (Data Acquisition and Distribution Service). Each element of requestable data (a subsystem or sensor status, or an engineering measurement) is uniquely identified by a measurement identifier (MSID).

One purpose of the present project is development of an interactive graphical display of the status of a simulation of the Environmental Control and Life Support subsystem (ECLSS). Sets of MSID's defined in the graphics software are transported over the network to the ECLSS simulation software using DADS. DADS then returns the simulated status of sensors aboard ECLSS, which are then displayed on a high-resolution computer graphics terminal. The software package is interactive, allowing the user to display the status of various components of the ECLSS subsystem.

A second, equally important, aspect of this research project is concerned with the impact of "standard" hardware and software on implementation of the software package. Developers of the DMS test bed have declared the MicroVax GPX as the standard display device. GKS is a well-known standard for rendering two-dimensional graphical images, which is available on DEC hardware. Finally, Ada has been adopted as the standard programming language aboard the Space Station. Accordingly, the software package was developed using a mixture of FORTRAN and Ada (for reasons to be discussed below). It is anticipated that lessons learned in this research can be helpful to those who design the final software aboard Space Station.

II. SOFTWARE DESIGN

Two features of the software design were judged especially important. First, the software should be designed in such a way that additional display capabilities could be added with a minimum of alteration to the original package. Second, the software package should be transportable across host processors equipped with GKS software.

The first consideration was accomplished by coding a control structure with "stubs" for added simulations as they become available. A control section written in Ada is essentially a menu-driven display processor, the highest level of which allows the user to select a subsystem for display. The display requirements for each subsystems can be made completely invisible to the main routine. Alternately, the display requirements can be met using a graphics parser written in FORTRAN and included as a package accessible by the main Ada supervisory routine. The graphics parser is general enough to provide all graphics services. Its capabilities can easily be extended as needed by extension of its dictionary of recognized graphics commands. Network access is provided by a separate Ada package which can be modified independently of the supervisory routines.

A second feature of the program design is its minimal reliance on "hard-coded" display items. We accomplished this by utilizing ASCII input files containing graphics commands directed to the graphics parser. Dynamic components of each display window are maintained as retained graphical segments which are defined once and for all at startup time and inserted into each display window as required. This strategy results in a flexible display manager: Static components of each display window can be changed by merely editing the corresponding

file with a text editor without recompiling the display program itself. Design of retained segments representing standard components of a schematic display, for example, make it easier for system-wide users to interpret graphical displays of system status.

The second consideration, software transportability, has been addressed by mixing languages in the demonstration program. The reasons for this are discussed in the next Section, where we consider the impact of the development environment on program design.

III THE IMPACT OF STANDARDS ON SOFTWARE DESIGN

GKS is itself merely a method for the device-independent management of two-dimensional graphics displays. As such it includes procedures for storage and manipulation of graphical segments, for implementation of software windows and view ports, and for interaction with graphical input devices such as the keyboard and mouse. Device independence allows the applications programmer to concentrate on the display task at hand, without the need to consider the specific characteristics of a particular hardware device. The functional con trol of graphics displays afforded by the GKS routines is judged adequate for the graphical manipulations required in the current display task.

The DEC/Vax implementation of the GKS standard has a significant impact on programmers writing in FORTRAN/77 or Ada. The GKS package used in this research is revision 3.0 of the DEC Vax software produced by Digital Equipment Corporation. The issues we consider in this Section are: (1) Existence of standard interfaces to GKS through FORTRAN and Ada; (2) Documentation and ease of software use by graphics novices and experts; (3) "Low-level" support provided for control of graphics input devices and display hardware; and (4) The availability of expert consulting services from the software vendor.

There are several interfaces to GKS procedures provided by DEC. Each programming language (Ada, Basic, C, FORTRAN, Pascal, etc.) uses a different interface, so that no single set of procedure calls exists. That is, the names of the routines are not the same in different languages, and the number and type of arguments passed to procedures differ from language to language. Here we focus on three such interfaces, Vax FORTRAN/77 to VAX GKS, ANSI FORTRAN/77 to GKS using the standard FORTRAN binding, and Vax Ada to Vax GKS. As of this writing there is no standard binding between Ada and GKS. Two such bindings have been proposed independently by ANS and ISO, but neither is available on DEC machines at present. (Parenthetically, we remark that an alternative interface to graphics routines is provided by UIS, the native graphics interface on the MicroVax. This is probably the highest performance, most powerful interface, but could not be used in this research. First, it is specific to the MicroVax GPX, and therefore requires non-portable code. Second, its use requires compilers and linkers which are unavailable on the DMS MicroVax. Third, DEC currently supplies no interface to UIS routines from Ada.)

The net result of the profusion of interfaces to GKS and their varying levels of documentation and functionality is that we chose to use a mixed-language program: The control structure was written in Ada, while direct access to graphics was provided by the FORTRAN/77 binding to GKS. We accomplished this by writing an Ada package specification containing pragma interfaces which allowed importing the FORTRAN/77 binding to GKS. Several procedures found to be only accessible through FORTRAN are accessed only through that language. The net result of this strategy is the generation of a single, standard interface between either FORTRAN/77 or Ada and the portable GKS binding. Transportability of the graphics software is thus possible

between processors equipped with FORTRAN/77, Ada, and GKS. Only a single level of documentation on the interface to GKS need be provided to FORTRAN and Ada programmers using this system.

We found the documentation provided by DEC to be generally inadequate. The highest level of documentation is provided for Vax FORTRAN/77 to Vax GKS. There is only minimal documentation on the FORTRAN/77 binding to GKS, and no documentation on the Vax Ada to Vax GKS interface. Several features of the FORTRAN/77 binding to GKS are completely unspecified in the system documentation, even though this interface is claimed to meet the highest standard of portability between processors equipped with GKS software.

This lack of adequate documentation is particularly undesirable. The complexity and profusion of procedure calls required to perform simple tasks is certain to confuse programmers with little training in graphics. Code fragments produced to document each procedure call are discussed briefly, but examples of only the simplest type of procedure call are given. The environment in which a given procedure call can be made is not generally discussed, so that one is required to read and study the entire two-volume manual set in order to trace errors. DEC has a rather unorthodox model of segment storage which can be confusing to the novice programmer. Its method of segment definition requires careful study, and is largely undocumented. Finally, access to graphics primitives (a circle, for example) is provided through graphics drawing primitives (GDP's). This is a most unorthodox method of generating these primitives, is tedious to code, largely undocumented, and error-prone. We have had extensive experience with a wide variety of graphics software and hardware. This implementation of GKS is by far the most complex one we have encountered. In fact, though

Ada was unknown to us at the start of the research project, it was much easier to learn that language than it was to become familiar with GKS.

Tasks requiring precision control of graphics input devices are not easily accomplished in the Vax implementation of GKS. We have been unable. for example, to find a means of controlling the foreground and background colors of a "pop-up" menu from software. The font used in displaying the menu items is not under software control, unless one is prepared to write a special-purpose graphics handler routine (outside of GKS). Control of hardware windows appears possible only by use of the non-standard Vax FORTRAN/77 to Vax GKS interface. Importantly, the GKS software does not make use of the specialized hardware aboard the MicroVax GPX. Highly graphics-intensive applications run unacceptably slowly on the MicroVax for this reason.

Finally, we find that software support from DEC appears to be largely inadequate. Those to whom we have spoken are largely unknowledgable of grahics in general, and of procedural inter-relationships in DEC's GKS in particular. There seems to be little support given to the standard FORTRAN/77 binding to GKS.

IV. CONCLUSIONS

Our experience gained with DEC hardware, the DEC implementation of the GKS graphics standard, and with Ada can be summarized as follows. DEC hardware seems adequate to the display tasks which form one part of the present research project. Highly graphics-intensive applications run slowly on the MicroVax GPX when it is programmed using the various DEC implementations of the GKS standard. Use of the GKS standard for graphics can provide code portability provided the FORTRAN/77 to GKS binding is used.

The lack of adequate documentation on the GKS software makes it unlikely that novice graphics programmers will be able to use the system easily. It is therefore likely that modern display techniques, which make use of multiple windows and view ports, and of segment transformation techniques, will remain largely the province of experienced graphics designers. This is unfortunate, as the use of modern techniques makes for quickly-developed, expandable, transportable computer codes.

The Ada programming language is easy to learn, and possesses many desirable features. It is strongly favored in recursive applications, such as the menu-driven display processor developed here. Since no binding to Ada is currently available it makes sense to provide a procedural interface between Vax Ada and the FORTRAN/77 binding. This was accomplished, and can be used by other generations of graphics programmers who must program in the development environment provided by DEC.

Finally, it appears that there are in fact no "standards" yet available on DEC hardware. Though GKS is a graphics standard suitable for two-dimensional display management, the only interface presented to meet a portability standard is provided by the largely undocumented binding to FORTRAN/77. No similar binding of GKS to Ada exists. Indeed, no link between Ada and the UIS procedures native to MicroVax yet exist. The net result is a lowest-common-denominator software implementation which is not designed to make most efficient use of existing hardware.

ATMOSPHERIC ENTRY HEATING OF COSMIC DUST

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1987

Johnson Space Center

Prepared by:	George J. Flynn, Ph.D.
Academic Rank:	Assistant Professor
University & Department:	State University of New York -- Plattsburgh Dept. of Physics Plattsburgh, NY 12901
NASA/JSC	
Directorate:	Space and Life Sciences
Division:	Solar System Exploration
Branch:	Experimental Planetology
JSC Colleague:	David McKay, Ph.D.
Date:	August 25, 1987
Contract Number:	NGT 44-001-800

ABSTRACT

A computer simulation of the atmospheric entry deceleration and heating for micrometeorites into a planetary atmosphere was developed. The results of this model were compared to an earlier model developed by Whipple and extended by Fraundorf. The major difference between the extent of heating experienced in the two models results from an underestimation of the atmospheric density at altitudes above 130 km in the earlier model. Thus the Whipple/Fraundorf model systematically overestimates the peak temperature reached on atmospheric entry. The discrepancies are small for near vertical entry and/or high density particles, where little deceleration is experienced at high altitudes. For particles entering at grazing incidence and/or of low density the discrepancies are more pronounced.

Gravitational enhancement, which is a function of geocentric velocity at the collection opportunity, was found to bias near earth cosmic dust collections in favor of low velocity particles. The effect is to increase the proportion of low velocity dust, predominately from asteroids, in the stratospheric cosmic dust collections and on earth orbiting spacecraft impact surfaces over its proportion in the interplanetary dust cloud. These collections, thus, do not represent an unbiased sample of the interplanetary dust. If, however, the velocity distribution of each particle type can be established the interplanetary abundances could be calculated knowing the near earth gravitational enhancement.

INTRODUCTION

Cosmic dust particles are micrometeorites which are sufficiently small (less than 100 micrometers in diameter) to enter the earth's atmosphere without melting. After deceleration by atmospheric drag, the particles descend into the stratosphere where they are concentrated because of their low settling rate in this region. They are collected from the stratosphere on small impact collection plates carried on U-2 and RB-57 aircraft in the NASA Cosmic Dust Sampling Program at the Johnson Space Center.

One catagory of collected dust is called "chondritic" because of the similarity of the major element abundances in these particles to those in the carbonaceous chondrite meteorites. Members of the chondritic class of cosmic dust have been shown to be extraterrestrial because of the presence of solar wind noble gases and solar flare radiation damage "tracks" in the particles. Although the major element abundances are generally similar to the primitive CI carbonaceous chondrites, enrichments in volatile elements above CI levels (van der Stap et.al., 1986) suggest the particles formed in a different region or at a different time in solar system evolution than the parent bodies of the CI meteorites. The chondritic particles, when examined on the microscale in the Transmission Electron Microscope (TEM), are seen to be aggregates of submicron or micron sized crystals (dominantly olivines, pyroxenes, or layer-lattice silicates) in an even finer grained carbon rich matrix. Examination of crystals within a given particle shows them to be hetrogeneous, non-equilibrium mixtures of minerals (Fraundorf, 1981). This suggests the particles are less metamorphosed than even the most primitive CI meteorites, and thus better sample the primitive solar nebula. It has even been suggested, principally on the basis of large Deuterium/Hydrogen fractionation, that the chondritic cosmic dust particles contain better preserved remnant interstellar grains than the most primitive meteorites (Clayton, 1986).

Larger samples of the same material would be valuable to constrain the process of solar system evolution. Thus one key objective in studying the cosmic dust is to determine its source or sources. The particles contain evidence of their exposure to the solar wind, which penetrates only a few hundred Angstroms into the surface, indicating they existed in space as small particles not much different from their size and shape as recovered from the stratosphere. These particles could not, however, have existed as small objects in space for the entire 4.5 billion years since the formation of the solar system. Poynting-Robertson drag, an interaction between the particles and solar radiation,

causes particles in this cosmic dust size range to spiral into the sun in times of 10^4 to 10^5 years. Thus sources, active in the recent past, are required to provide the cosmic dust now being collected.

These sources, because of the primitive material which they contain, would be suitable targets for sample return missions. However the sources have not yet been identified. One of the major objectives of the Cosmic Dust Collection Facility, proposed for the Space Station, is to determine the velocity vector of each cosmic dust sample collected, and thus allow particles to be traced back to their individual sources. Prior to the launch of the Space Station, cosmic dust sources can only be inferred from the properties of the particles collected from the stratosphere.

In principle, almost every solar system object is able to contribute to the cosmic dust enviornment through outgassing, ejection due to cratering events, collisional fragmentation, tidal disruption, or volcanic activity. However the Infrared Astronomy Satellite (IRAS) has detected two major sources of dust in the solar system: the main asteroid belt and comets.

Particles from these sources can only be collected at earth when their orbits intersect the orbit of the earth. Generally the source orbits are not earth crossing. The orbit of the particle must then evolve from that of the parent body to an orbit which is earth intersecting. The dominant force causing this orbital evolution is Poynting-Robertson (PR) drag. Flynn (1986, 1987) has shown that, under the influence of PR drag particles arriving at earth from asteroidal and cometary sources differ significantly enough in their earth encounter geometry that they can be distinguished on the basis of the magnitude of their encounter velocity (though the full velocity vector, as will be measured on the Space Station Cosmic Dust Collection Facility, will be necessary to distinguish individual sources within the general catagories).

The major effect of PR drag is to decrease the aphelion of an initially elliptical orbit with little change in the perihelion until the orbit is near circular. Once the orbit is nearly circular the particle spirals into the sun. This effect is illustrated in Figure 1 for a particle released into the orbit of Comet Encke.

Particles originating in the main asteroid belt are initially in near circular orbits of relatively low inclination to the the plane of the earth's orbit. Under PR drag they spiral in towards the sun. The earth collection opportunity occurs when the particle orbit has a radius of about 1 AU. Thus collection is from a near circular orbit of low inclination, and the particle has a very low geocentric velocity at the collection opportunity.

Particles derived from comets divide naturally into two catagories. Those from comets with perihelia significantly larger than 1 AU have evolved to near circular orbits before the perihelion has decreased to 1 AU. Thus collection again occurs from a near circular orbit, giving a low geocentric velocity. However comet orbits are generally more highly inclined to the earth's orbital plane than are main belt asteroidal orbits. Thus collection occurs from an inclined orbit, resulting in a higher geocentric velocity at collection than for the main belt asteroidal case.

Particles from comets with smaller perihelia are collected from orbits which are still elliptical at the collection opportunity, thus giving rise to an even higher geocentric velocity at the collection opportunity. If the initial comet orbit is retrograde, as is the case for Comet Halley, the geocentric velocity at collection is still higher. The results, for main belt asteroids and a variety of comets, are given in Table I.

Since the heating experienced by each particle depends on its atmospheric entry velocity, particles from these three groups should experience different degrees of entry heating. Thus atmospheric entry heating may provide a suitable criteria to distinguish asterioidal from cometary sources.

Atmospheric Entry Heating

The first detailed model of atmospheric entry heating for micrometeorites was developed by Whipple (1950; 1951). He demonstrated that the peak temperature reached on atmospheric entry depends on the particle velocity, the angle between the normal to the earth's surface and the velocity vector at entry (with particles making more grazing entries being substantially less heated), as well as properties of the atmosphere and the particle itself. Fraundorf (1980), expanding on the Whipple model, derived a closed form solution for the probability that a particle would be heated above an arbitrary temperature T on entry given an entry velocity and assuming a random distribution of entry angles. Using the Fraundorf solution to the Whipple entry heating model Flynn (1987) showed that the entry temperature range from 500^{o}C to 800^{o}C is critical in distinguishing asteroidal from cometary materials. This is a temperature range in which pulse heating simulations of atmospheric entry heating confirm that many particle alterations occur. Among them are the loss of solar flare tracks through annealing, alteration to the structure of minerals (particularly layer-lattice silicates), and loss of volatile elements. These internal thermometers suggest that a large fraction of the chondritic particles recovered from

the stratosphere derive from asteroidal sources. If true, this indicates a population of main belt asteroids significantly more primitive than sampled by the known meteorites. Indeed, the porous structure of the chondritic particles strongly indicates that larger meteors composed of this material would most likely fragment on entry, thus precluding recovery of larger samples among the meteorites.

If main belt asteroids are the sources of these primitive cosmic dust particles, than the asteroids may be as suitable as targets for primitive material sample return missions as the comets. However the distinction between asteroidal and large perihelion cometary sources rests on only a 100 to 200^{o}C difference in the peak temperature reached on atmospheric entry. The purpose of this study was to review the Whipple/Fraundorf entry heating model, assess the uncertainties in its peak temperature predictions, and, where possible, improve upon the earlier model.

Entry Heating Model

The process of atmospheric deceleration and heating of micrometeorites was simply modeled by Whipple (1950). In this model the incoming dust particle (moving at somewhere in the range from 11 km/s, a lower limit imposed by gravitational infall acceleration, to 70 km/s, the upper limit for confinment in the solar system) is thought of as participating in single particle colisions with the air molecules. A dust particle moving at a velocity v and having a cross sectional area A will sweep out a volume of air V in a time interval dt which is given by:

$$V = A \cdot v \cdot dt \qquad \text{(Equation 1)}$$

If the air density at a particular height, h, is given by a function p(h), then the total mass of air encountered in the time interval dt is given by:

$$M = p(h) \cdot V = p(h) \cdot A \cdot v \cdot dt \qquad \text{(Equation 2)}$$

Although the individual air molecules are moving, on the average their velocity and momentum are zero. Interaction with the incoming particle alters the net momentum of the air molecules. At one extreme, a totally inelastic collision, the air is hit by the dust particle, sticks, and moves along with the particle, undergoing a net momentum change of M·v. At the other extreme, an elastic collision, the air molecule, initially approaching the dust particle with a velocity v in the rest frame of the particle, bounces off with a velocity v in the opposite direction. In this case the net change of momentum of the air molecule is

$2 \cdot M \cdot v$. Thus the air molecules gain (or the dust particle loses) an amount of momentum dP in the interval dt given by:

$$dP = L \cdot M \cdot v = L \cdot [p(h) \cdot A \cdot v \cdot dt] \cdot v \quad \text{(Equation 3)}$$

where L varies from 1 to 2 as the collision ranges from perfectly inelastic to elastic.

Since the particle must undergo a momentum loss of the same magnitude, the particle slows by an amount dv, given by:

$$dv = dP/m = L \cdot [p(h) \cdot A \cdot v \cdot dt] \cdot v \quad \text{(Equation 4)}$$

where m is the mass of the dust particle. Equation 4 gives the deceleration of the particle (dv/dt) for any height and velocity provided the variation of air density with height is known. With this deceleration the variation of the velocity with height can be calculated.

During the deceleration the dust particle will heat up if the collision process is not elastic. In an inelastic collision the excess kinetic energy is transformed into heat, which can be used to warm the particle or can be radiated away. The amount of heat lost, H, in the time interval dt, which depends on the particle temperature T, the surface area S, the emissivity e, and the Stefan-Boltzman constant b, is given by:

$$H = b \cdot e \cdot S \cdot T^4 \quad \text{(Equation 5)}$$

In that same time interval, if the particle has warmed or cooled from T_o at the start of the interval to T at the end of the interval an amount of heat, Q, must have been added or taken from the particle. If the particle has a specific heat C, the Q is given by:

$$Q = m \cdot C \cdot (T - T_o) \quad \text{(Equation 6)}$$

The energy gained by the particle, E, in the time interval dt is some fraction, K, of the total kinetic energy of the air molecules. K must, then, vary from 0 for the elastic case to 1 for the totally inelastic case. Since the kinetic energy of the air molecules is given by $0.5 \cdot M \cdot v^2$, the energy gained by the particle is

$$E = 0.5 \cdot K \cdot M \cdot v^2 \quad \text{(Equation 7)}$$

The temperature required for the dynamic equilibrium in which energy is radiated away as fast as it is added in the time interval is then obtained by equating the energy input E to the sum of the radiation loss H and the heat energy Q. This gives:

$$0.5 \cdot K \cdot M \cdot v^2 = b \cdot e \cdot S \cdot T^4 + m \cdot C \cdot (T - T_o) \quad \text{(Equation 8)}$$

Since M is a function of height, as given by Equation 2, and v can be obtained as a function of height from the deceleration derived in Equation 4, Equation 8 can be solved to give the particle temperature as a function of height (or time). However, since M involves the atmospheric density p(h), Whipple found that Equation 8 could easily be solved in closed form with two assumptions:

1) the specific heat term is negligible, and can be ignored, and,

2) the atmosphere is isothermal, and of constant composition, so that p(h) varies exponentially with height.

Fraundorf followed these assumptions of Whipple to derive the probability that any given particle arriving at earth with a random orientation but specified velocity would be heated above a temperature T.

In order to apply the Whipple model values of the drag parameter, L, the accomodation coefficient, K, and the emissivity, e, had to be assumed. Following the earlier work, these were all taken to be 1. In order to assess the validity of the model each of these assumptions was considered separately.

Drag Parameter

Whipple argued that the collisions were likely to be very close to pure inelastic collisions since the energy of each air molecule in the dust particle's reference frame is tens to hundreds of eV. Thus air molecules will penetrate several atomic layers into the particle before stopping. Although these air molecules may eventually boil off as the particle heats up, the collision process is essentially inelastic.

An alternative way to look at the interaction of the dust particle with the air is in the context of normal aerodynamic drag. In this context, the drag force, F, is given in terms of the drag coefficient, D, of the particle by:

$$F = 0.5 \cdot D \cdot A \cdot p(h) \cdot v^2 \quad \text{(Equation 9)}$$

Since the drag force can be rewritten as the time rate of change of the momentum (dP/dt), the momentum loss dP in the time interval dt is given by:

$$dP = 0.5 \cdot D \cdot A \cdot p(h) \cdot v^2 \cdot dt \quad \text{(Equation 10)}$$

Equation 10 is identical to Equation 3 when the drag parameter L is identified as 0.5 times the drag coefficient.

Although drag coefficients have not been measured for objects in the cosmic dust size range in air densities comparable to the upper atmosphere, they have been derived for orbiting satellites on the basis of the rate of their orbital decay. King-Hele (1964) argues that for satellites of a variety of shapes, ranging from spheres to tumbling cylinders, the drag coefficient can be taken to be equal to $2.2 \pm 5\%$. This implies a value of the drag parameter L equal to 1.1, a 10% increase from the assumed value of 1. The effect of increasing the drag parameter by 10% is to decrease the maximum temperature on entry by 2.5%, an insubstantial correction. Thus the value of 1 assumed by Fraundorf seems not to be in serious error.

Accomodation Coefficient

If the value of the drag parameter is taken as near 1, then the accomodation coefficient cannot dramatically differ from 1. The two are related in that if the air molecules stick to the particle surface then they must transfer all their kinetic energy to the particle.

Emissivity

For the emissivity a value of 1, indicating a perfect black body, was adopted by Whipple and Fraundorf. Optical microscope observation of the cosmic dust recovered from the stratosphere indicates a black color for many particles, consistent with a high absorptivity (and thus emissivity) in the visible region of the spectrum. In the infrared, absorption spectra indicate a substantial fraction of the incident light is absorbed at all wavelengths. While the results have not been quantified by the investigators in the IR absorption experiments, the results are consistent with high IR absorption even on particles crushed and spread out in order to reduce their thickness. The peak of the black body emission curve for objects in the 1000 K to 1500 K varies from 2 to 3 micrometers, substantially smaller than the typical particle dimensions of 10 to 25 micrometers. While there is not a direct measurement of the emissivity of these particles in the IR region, the available data is consistent with an emissivity not substantially below 1. Reducing the emissivity to 0.8 would increase the peak temperature on entry by 5.7%, again a negligible effect.

Atmospheric Density

Whipple assumed an exponential decrease of mass with height, consistent with an isothermal atmosphere of constant composition, in the absence of good experimental determinations. Fraundorf followed this approach because it permits a closed form solution to the entry heating equations. However, since Whipple's 1950 model, a substantial body of data on the density of the upper

atmosphere has become available. A comparison of the U.S. Standard Atmospheres (1962) density with the exponential approximation (Figure 2) shows reasonable agreement up to about 130 km. Above that the U.S. Standard Atmosphere is significantly more dense, and thus has more stopping effect, than the exponential model.

To assess the magnitude of this effect a computer simulation of the entry heating was performed. In this simulation the entry heating dynamics proceed as described for the Whipple/Fraundorf model, except that at every height the atmospheric density is read from the U.S. Standard Atmospheres data table rather than using the exponential approximation. During the time when the particle is experiencing significant heating a 2/100 sec time intrerval was employed in the simulation.

Under the effect of gravitational infall, the velocity profile of a particle starting from 2×10^8 meters with a velocity of 10 km/s was calculated in the absence of an atmosphere. The same velocity profile was then calculated for a spherical dust particle of 20 micrometer diameter and density 1 gm/cm^3 encountering the U.S. Standard Atmospheres density profile. The results, shown in Figure 3, show that the dust particle reaches its maximum velocity at 176 km above the surface, indicating that it is already experiencing significant deceleration at that height. At that time the atmospheric density is an order of magnitude higher in the U.S. Standard Atmospheres table than in the exponential approximation. This indicates the exponential approximation may be a source of significant overestimation of the peak temperature reached on entry.

A direct comparison of the peak temperatures reached on normal incidence entry gives 1185 K in the Whipple/Fraundorf model versus 1159 K in this model. The effect appears to be small, however it difference between the two approaches would be expected to increase for grazing entry conditions, when the particles spend a longer time in the outer regions of the atmosphere.

To provide a direct comparison, the computer simulation was modified to allow the impact parameter (the distance of the incoming particle from a line through the center of the earth and parallel to the particle trajectory), d, of the incoming dust particle to be varied. Varying the impact parameter is the equivalent of assuming a random distribution of velocity vector orientations with respect to the top of the atmosphere. The impact parameter was varied in uniform 0.5×10^5 meter steps from zero until earth collection was no longer possible. (This maximum impact parameter varies with the velocity of the incoming particle as will be described in the section on Gravitational Focusing).

The fraction of particles not heated above a temperature T was calculated using both the computer simulation and the Fraundorf equations. The results, shown in Table II, indicate that the most serious discrepancies between the two models are at grazing incidence (which corresponds to low peak temperatures). The computer simulation gives an increase by a factor of 20, from 0.1% to 2%, in the fraction not heated above 236^{o}C for example. Similar large differences are seen between the models for particles of very low density, since these particles experience considerable deceleration in the outer regions of the atmosphere, where the air density is significantly underestimated by the exponential model.

The peak temperature versus impact parameter is shown in Figure 4 for a 20 micrometer diameter, density 1 gm/cm^{3} particle. esults are shown for three different starting velocities 1 km/s, 10 km/s, and 20 km/s. If 1200 K is taken as the melting temperature for the lowest temperature minerals in the particles, then virtually all particles with a geocentric velocity (before gravitational infall acceleration) of 20 km/s would be expected to show signs of melting. The temperature for solar flare track annealing is reported as 800 to 900 K (Fraundorf et.al., 1982), indicating that virtually all particles of this size and density with a geocentric velocity of 10 km/s would lose their solar flare tracks on entry. The presence of solar flare tracks in a large fraction of the stratospheric cosmic dust particles thus indicates a rather low geocentric velocity (no more than a few km/s) at collection. This is consistent with an asteroidal source for the majority of the particles collected at earth.

Gravitational Focusing

Perhaps the most striking feature of Figure 4 is the dramatic increase in the maximum impact parameter from which earth collection is possible as the geocentric velocity decreases. The earth effectively presents a bigger target for low velocity particles than for particles of higher velocity. Opik (1951) pointed out that the cross-section for collection of a small particle by a larger body varies with the relative velocity, v, between the two particles and the escape velocity, v_e, from the large object. This cross-section is given by:

$$CR = 3.14 \cdot R^2 (1 + v_e^2/v^2) \quad \text{(Equation 11)}$$

where R is the radius of the larger object. For the case of cosmic dust being collected at or near earth from the asteroidal and cometary sources previously discussed, the

effective cross-sections are given in Table I. The gravitational enhancements, that is the ratios of the effective cross-sections to the geometrical cross section, range from 126 for 1 km/s asteroidal dust to 1 for the highest velocity cometary dust from Halley.

This near earth gravitational enhancement indicates that if any particles with low geocentric velocities exist among the interplanetary dust, the proportions of those particles will be substantially enhanced relative to higher geocentric velocity particles in any near earth collection process. Thus the near earth collectors (stratospheric or satellite) do not provide an unbiased sample of the interplanetary dust cloud. Near earth collection is heavily biased towards the low geocentric velocity fraction of the interplanetary dust.

The actual distribution of velocities for particles in the cosmic dust size range has never been measured. These particles are not sufficiently heated on entry to produce luminous or ionized trails, which are the basis for detection of visual or radar meteors. The radar meteors, which are the smallest size meteors routinely detected, begin at about 100 micrometer diameter and range upwards in size. These meteors have a mean geocentric velocity (before gravitational infall) of 9 km/sec (Southworth and Sekanina, 1973). However there is reason to believe that particles smaller than 100 micrometers in diameter may differ in source, and thus in velocity distribution, from the larger micrometeorites (Flynn, 1987; Zook and McKay, 1986). If so, the radar meteor data for larger micrometeorites would not be indicative of the distribution for particles in the cosmic dust size range.

The observation of microcraters, in the size range produced by cosmic dust particles, on exposed lunar rock surfaces indicates that a large majority of the incoming particles in this size range have sufficient velocity to produce craters. However simulations indicate that a dust velocity of 3 km/s is sufficient to produce glass lined microcraters (Ashworth, 1978). This is not significantly higher than the earth-moon system escape velocity from the lunar surface, thus it does not significantly constrain the in space velocity of the cosmic dust particles.

The degree of near earth enhancement of the micrometeorite flux may provide the most significant constraint on the abundance of low geocentric velocity dust in the interplanetary dust cloud. Micrometeorite penetration detectors of similar design were flown on two earth orbiting satellites, Explorer XVI and Explorer XXIII, and on the five Lunar Orbiters. Each satellite carried pressurized-cell penetration detectors with 25 micrometer thick walls, thus measuring particles in the cosmic dust size range. The penetration rates, corrected for satellite and earth shielding, were 0.445 events/$m^2 \cdot$day and 0.526 events/$m^2 \cdot$day

respectively (Naumann, 1966). The corresponding shielding corrected penetration rate for the five Lunar Orbiters was 0.19 events/m^2·day (Grew and Gurtler, 1971). These penetration rates imply a near earth enhancement over the lunar flux of 2.3 to 2.8. When corrected for the small enhancement at the moon, due to its own gravitational field, the earth enhancement over the interplanetary flux at 1 AU would be from 2.5 to 3.0.

There is considerable uncertainty in the near earth enhancement factor because of the small number of penetration events (only 22 for the five Lunar Orbiters) detected over the operational lifetimes of the satellites. In addition, the reported lunar flux may be too high due to secondary debris from lunar surface impacts striking the Lunar Orbiter spacecraft. However, taking 2.8 as the current best estimate of the near earth enhancement does apply some constraints to the velocity distribution, and thus the possible sources, of the cosmic dust.

This observed near earth enhancement clearly excludes a micrometeorite population composed solely of particles from cometary sources with samll perihelia. As seen in Table I, these sources all supply particles with enhancement factors below 2. This enhancement also appears to exclude a completely asteroidal population, since a minimum enhancement of 6 would be expected in that case. However, recent results by Gufstafson and Misconi (1986) suggest that earth scattering of asteroidal particles before the collection opportunity may increase the inclinations and ellipticities of their orbits substantially above those calculated by PR drag alone. Thus the geocentric velocities of the asteroidal particles at the collection opportunity are presently somewhat uncertain. This should improve as the process of earth scattering is better simulated.

A single source model is unrealistic in any case, since the IRAS measurements demonstrate that both comets and main belt asteroids make significant contributions to the interplanetary dust. We can place some limits on the proportions of dust from low and high velocity sources by requiring that the overall gravitational enhancement equal the observed value of 2.8. If we consider the dust to be composed of two components, one with a relatively high geocentric velocity, and the second with lower geocentric velocity, we can calculate the proportions of dust which would be observed on near earth collectors for any given interplanetary mixture. For the high velocity component I have taken dust with an enhancement factor of 1.2, the value appropriate for dust from Comet Encke.

If the interplanetary dust cloud were to consist of 99% material from this source and 1% material from a low velocity source, then an enhancement factor of 160 would be required for the low velocity component in order to produce

the observed overall enhancement of 2.8. If the low velocity source had a single velocity, a velocity of 0.9 km/s would be required to produce this enhancement. In this case, although the space proportions were taken as 99% high velocity dust and 1% low velocity dust, near earth collectors would see a flux of 57% low velocity and 43% high velocity dust because of the dramatic near earth enhancement of the low velocity component. The results for a variety of other interplanetary mixtures are given in Table III.

The extent of the near earth collection bias is clearly seen in Table III. If the high velocity component has an enhancement of only 1.2, then to produce the observed overall enhancement of 2.8 results in a near earth collection dominated by low velocity dust for any interplanetary mixture of the two components. If the high velocity component is assumed to have a higher enhancement, such as the 1.7 for d'Arrest, the required low velocity enhancement is decreased. However the fraction of low velocity material collected near earth is still dramatically enhanced over the interplanetary proportions.

Near earth cosmic dust collections are dramatically biased towards the collection of the low velocity component of the interplanetary dust cloud. Detailed measurements on the velocity distribution of the near earth flux will allow the extent of the bias to be evaluated.

Conclusions

The atmospheric entry heating model developed by Whipple and extended by Fraundorf accurately assesses the entry heating of cosmic dust particles of moderate or higher density and for incidence angles near the normal. As the particle density decreases or the entry angle approaches a grazing condition the Whipple/Fraundorf model systematically overestimates the peak temperature reached on entry since there is then substantial deceleration in the upper region of the atmosphere, a region where their use of an exponentially decreasing atmospheric density underestimates the deceleration. For particles of density 1 gm/cm^3 the differences between this simulation and the closed form solution of Whipple and Fraundorf are not significant enough to alter the conclusion in Flynn (1987) that, based on the heating experienced on atmospheric entry, the majority of the chondritic cosmic dust collected from the stratosphere encountered the earth with a low geocentric velocity. This is consistent with an asteroidal origin for the chondritic cosmic dust, and distinctly different from what would be expected for dust derived from comets with small perihelia (such as Eecke) and evolving into earth intersecting orbits under the Poynting-Robertson drag force.

Gravitational enhancement of the flux of low velocity dust produces a bias in near earth collections. The effect is to enhance the proportions of asteroidal dust, and to a lesser extent dust from large perihelion comets, relative to that of dust from the large perihelion comet population in interplanetary dust cloud. Thus, if main belt asteroids contribute at all to the the interplanetary dust cloud, the expected low geocentric velocity of dust from this source would produce substantial enrichment of this material in the stratospheric and satellite collectors. The cosmic dust collected from the stratosphere or on impact surfaces of earth orbiting spacecraft does not represent an unbiased sample of the material making up the interplanetary dust cloud. The true proportions of the interplanetary cloud material can be inferred from the near earth collections if the velocity distribution, and thus the gravitational enhancement, of each particle type is established.

Table I

Relative Velocity at Earth Collection and Earth Enhancement
For Dust From Asteroidal and Cometary Sources
(Particle: Density 1 gm/cm^3, Diameter 20 micrometers)

Parent Object Orbit	Geocentric Velocity* (km/s)	Near-Earth Enhancement**
Main Belt Asteroid		
Circular, 0° Inclination	1	126
Circular, 6° Inclination	3	15
Circular, 10° Inclination	5	6
Comets (perihelia > 1.2 AU)		
Giacobini-Zinner	19	1.3
d'Arrest	13	1.7
Halley	64	1.0
Swift-Tuttle	51	1.0
Comets (perihelia < 1.2 AU)		
Kopff	6	4.4
Temple I	8	2.9
Temple II	9	2.5

* Gravitational infall velocity not included.
** Ratio of earth collection cross-section to geometric cross-section.

Table II

Fraction of Particles Not Heated Above Temperature T
(Particle: Density 1 gm/cm^3, Diameter 20 micrometers, Velocity 20 km/s)

Temperature (°C)	Fraundorf (1980)	This Model
236	0.1%	2%
406	1%	4%
471	2%	6%
545	5%	10%
615	10%	16%
680	17%	26%
781	39%	50%
841	57%	75%
870	75%	90%

Table III

NEAR EARTH ENHANCEMENT

High Velocity Fraction*	.99	.90	.75	.50	.25	.10	.01
Low Velocity Fraction	.01	.10	.25	.50	.75	.90	.99
Near-Earth L/H**	57/43	61/39	68/32	79/21	90/10	96/4	99/1
Low Velocity Enhancement	160	17	7.6	4.4	3.3	3.0	2.8
Charactaristic Velocity (km/s)	0.9	2.8	4.4	6.1	7.3	8.0	8.3

* The High Velocity fraction has an assumed enchancement factor of 1.2, corresponding to Encke dust.

** Collected Proportions of Low Velocity/High Velocity Dust

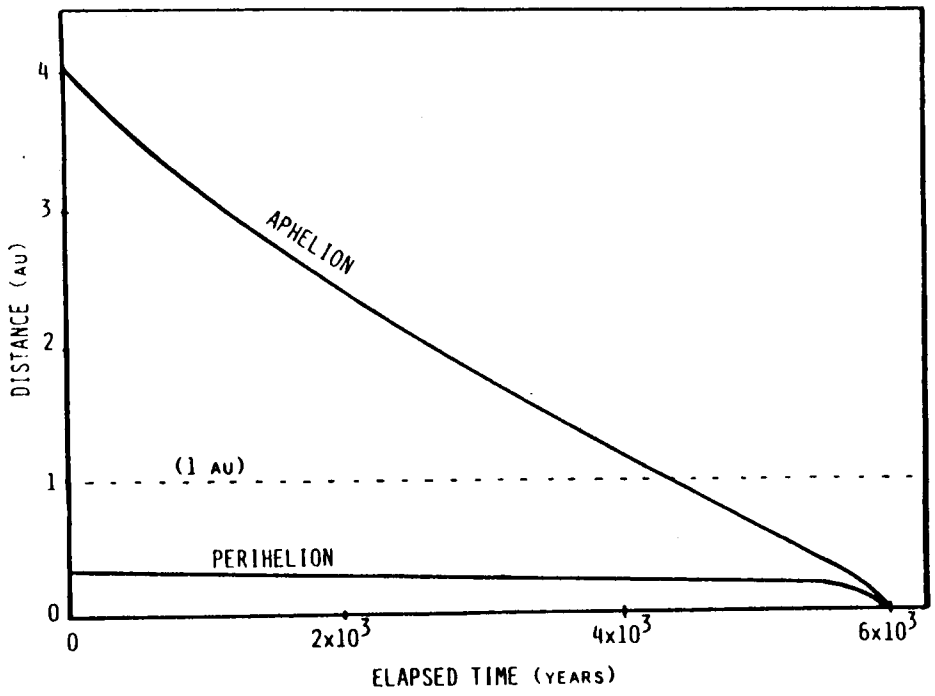


FIGURE 1. Orbital evolution for a 20 μm diameter, density 1 gm/cm^3 dust particle released into the orbit of Comet Encke under the influence of Poynting-Robertson drag.

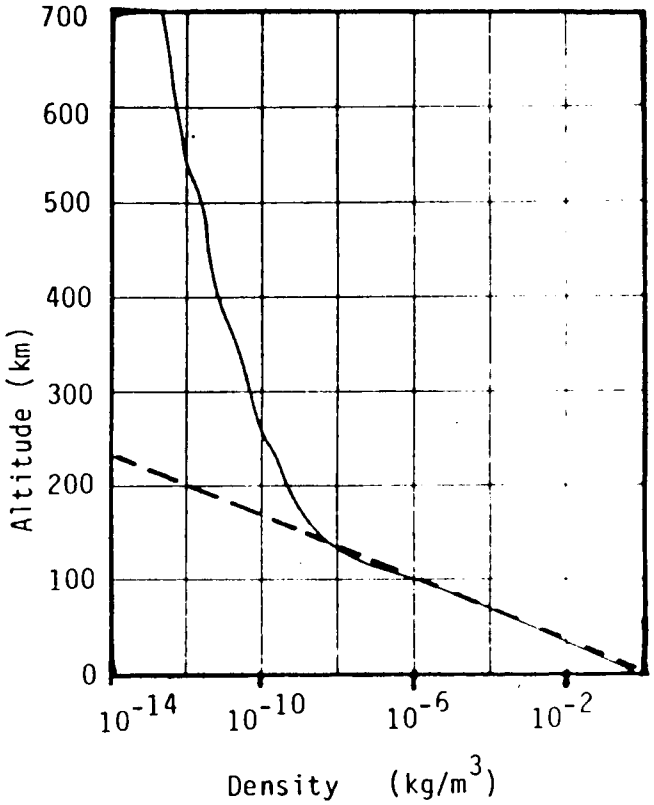


Figure 2. Comparison of the atmospheric density variation with altitude for the U.S. Standard Atmosphere (1962) (solid line) and the exponential approximation (dashed line).

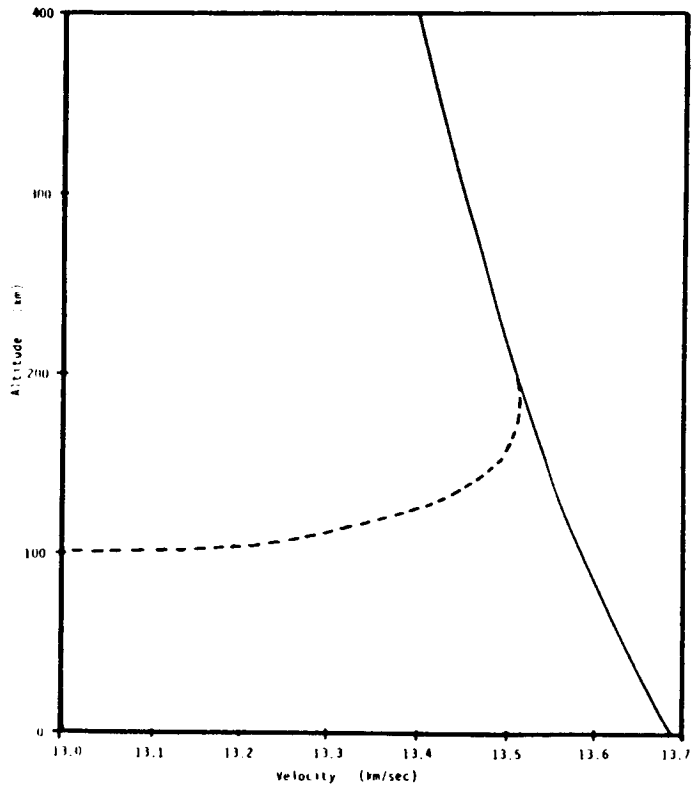


Figure 3. Velocity versus altitude for a particle released from infinity with a velocity of 10 km/s towards earth (solid line) and for the same particle encountering the U.S. Standard Atmosphere(1962) atmosphere. Particle: diameter 20 μm, density 1 gm/cm^3.

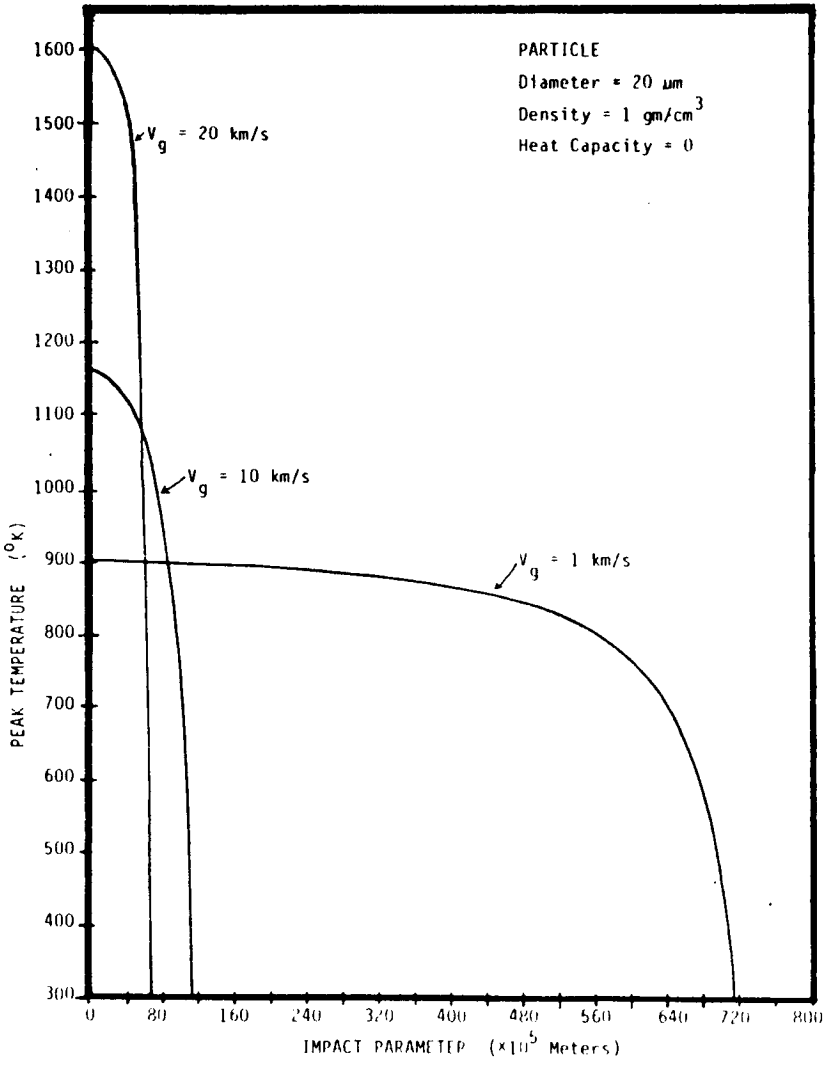


Figure 4. Peak temperature as a function of impact parameter for particles having velocities of 20 km/s, 10 km/s, and 1 km/s towards earth at 6 x 10^8 meters from earth.

REFERENCES

Ashworth, D. G., Lunar and Planetary Impact Erosion. In *Cosmic Dust*, ed. J. A. M. McDonnell, Wiley, Chinchester, p. 427-526, 1978.

Clayton, D. E., Interrelationships Between Interstellar and Interplanetary Grains. In *Interrelationships Between Circumstellar, Interstellar and Interplanetary Dust*, ed. J. A. Nuth and R. E. Stencel, NASA CP-2403, p. WG85-120, 1986.

Flynn, G. J., Effects of Orbital Evolution on Encounter Velocity and Exposure Age of IDP's, *Meteoritics*, Vol. 21, #4, p. 362-363, 1986,

Flynn, G. J., Atmospheric Entry Heating: A Criteria to Distinguish Between Asteroidal and Cometary Sources of Interplanetary Dust. Submitted to *Icarus*, 1987.

Fraundorf, P., The Distribution of Temperature Maxima for Micrometeorites in the Earth's Atmosphere Without Melting. *Geophys. Res. Lett., 10*, 765-768, 1980.

Fraundorf, P., Interplanetary Dust in the Transmission Electron Microscope: Diverse Materials from the Early Solar System. *Geochemica. Cosmochem. Acta, 45*, 915-943, 1981.

Fraundorf, P. Lyons, T., and Schubert, P., The Survival of Solar Flare Tracks in Interplanetary Silicates on Deceleration in the Earths Atmosphere. *Journ. Geophys. Res., 87*, A409-412, 1982.

Grew, G. W. and Gurtler, C. A., The Lunar Orbiter Meteoroid Experuiments, NASA TN-D-6266, 1971.

Gustafson, B. A. S. and Misconi, N. Y., Interplanetary Dust Dynamics I: Long-Term Gravitational Effects of the Inner Planets on Zodical Dust, *Icarus, 66*, 280-287, 1986.

King-Hele, D., Theory of Satellite Orbits in an Atmosphere, Butterworths, London, 1964.

Naumann, R. J., The Near Earth Meteoroid Enviornment, NASA TN D-3717, 1966.

Opik, E. J., Collision Probabilities With the Planets and the Distribution of Interplanetary Matter, *Proc. Roy. Irish Acad., 54, Sect. A.*, 165-199, 1951.

Southworth, R. B., and Sekanina, Z., Physical and Dynamical Studies of Meteors, NASA CR-2316, 1973.

U. S. Standard Atmospheres, 1962, U. S. Govt. Printing Office, 1962.

van der Stap, C. C. A. H., Vis, R. D., and Verheul, H., Interplanetary Dust: Arguements in Favor of a Late Stage Nebular Origin, *LPSC XVII*, 1013-1014, 1986.

Whipple, F. L., The Theory of Micro-Meteorites. Part I: In an Isothermal Atmosphere. *Proc. Nat. Acad. Sci., 36*, No. 12, 687-695, 1950.

Whipple, F. L., The Theory of Micro-Meteorites. Part II: In Hydrothermal Atmospheres, *Proc. Nat. Acad. Sci., 37*, No. 1, 19-31, 1951.

Zook, H. A., and McKay, D. S., On the Asteroidal Component of Cosmic Dust, *LPSC XVII*, 977-978, 1986.

ACTIVE CONTROL OF FLEXURAL VIBRATIONS IN BEAMS

Final Report

NASA/ASEE Summer Faculty Fellowship Program-1987

Johnson Space Center

Prepared by:	Carl H. Gerhold, Ph.D.
Academic Rank:	Associate Professor
University & Department:	Texas A & M University Mechanical Engineering Department College Station, Texas 77843
NASA/JSC	
Directorate:	Engineering
Division:	Structures and Mechanics
Branch:	Loads and Structural Dynamics
JSC Colleague:	A. Rodney Rocha
Date:	August 7, 1987
Contract Number:	NGT 44-001-800

ABSTRACT

The feasibility of using piezoelectric actuators to control the flexural oscillations of large structures in space is investigated. Flexural oscillations are excited by impulsive loads such as Shuttle docking. The vibratory response can degrade the pointing accuracy of cameras and antennae, and can cause high stresses at structural node points. Piezoelectric actuators have the advantage of exerting localized bending moments. In this way, vibration is controlled without exciting rigid body modes in the structure. The actuators are used in collocated sensor/driver pairs to form a feedback control system. The sensor produces a voltage that is proportional to the dynamic stress at the sensor location, and the driver produces a force that is proportional to the voltage applied to it. The analog control system amplifies and phase shifts the sensor signal to produce the voltage signal that is applied to the driver. The feedback control system is demonstrated to increase the first mode damping in a cantilever beam by up to 100 percent, depending on the amplifier gain. An analytical model of the control system has been developed. The estimated and measured vibration control compares favorably. A simulated free-free beam has been fabricated and instrumented with a distribution of piezoelectric sensor/drivers. The purpose is to evaluate the damping efficiency of the control system when the piezoelectrics are not optimally positioned at points of high stress in the beam. The control system is found to reduce the overall vibration response to impact by a factor of two.

INTRODUCTION

Oscillations in the large, flexible structures in space are induced by impact such as docking operations. Flexural vibration transmitted through the structure can degrade the pointing accuracy of devices such as cameras. Transmitted vibrations can degrade the microgravity environment in modules containing vibration sensitive equipment. Large oscillations in the structure can cause high stresses to occur within the structure and at node points. In the extreme case, flexural vibration can cause the structure to go into a rigid body tumbling mode. The oscillatory response of the structure can be controlled through material damping built into the structural elements, or by increasing the mass of the structural elements. However, addition either of mass or damping material imposes a weight penalty, which increases construction cost because of the added weight to be delivered into space.

It is proposed to develop an active control methodology to limit the vibratory response of a large, flexible structure. The purpose of the research conducted in the Summer Faculty Fellowship program is to establish the suitability of piezoelectric actuators to control bending in flexible structures. The piezoelectric actuators are used in pairs, bonded at opposite and equal distances from the neutral axis of the elastic member. When the elastic member is excited by a transient force, the free vibration causes bending moments which produce local stresses along the member's surface. One element senses the local stress and produces a proportional voltage. The voltage signal is amplified and shifted in phase by 180°.

This conditioned signal drives the companion piezoelectric element to produce a localized moment in the elastic member to oppose the dynamic bending. Piezoelectric actuators apply moments locally. Thus, they can control oscillations without exciting rigid body modes in the elastic member, as may be produced by reaction jets.

Piezoelectric actuators have been demonstrated to produce significant flexural control in cantilevered beams, where the sensor/driver pair are bonded near the root of the beam. It is known that the bending stresses are maximum near the root in all modes of vibration. Thus, the control is expected to be optimum for the sensor/driver pair mounted at this known point of high stress. In the current project, a distribution of piezoelectric element pairs is bonded on a free-free beam and the vibration control efficiency is evaluated. The element locations do not necessarily coincide with positions of highest bending stress in the beam. The goal is to determine the efficiency of the piezoelectric actuators when they are not optimally placed at locations of maximum bending stress. This configuration is chosen because it may not be possible to estimate accurately the bending modes of a structure with many substructures and appendages, such as a space structure. This is particularly true since the design is expected to be dynamic with substructure modifications made throughout its mission duration. The travel of the remote manipulator arm trolley also means that the mass distribution of the structure can be changing continuously, even between design changes.

When the sensor/driver pairs are collocated, that is equidistant from the neutral axis, the bending moment exerted by the driver is directly proportional to the stress at the sensor. Thus, the feedback control system is independent of the mode shape of the elastic member. The only expected limitation of the sensor/driver pair to control vibration in the beam is the amplification required between the sensor and driver. Tests have shown that the sensor voltage must be amplified by a factor on the order of 150 to control vibrations in a cantilever beam. Theoretically, any magnitude of amplification is achievable. Higher orders of gain may be required to control oscillations when the sensor/driver pair is not located at the positions of high stress in the beam where the voltage produced by the sensor is high. However, it is also found that the maximum voltage that can be applied to drive the piezoelectric element is approximately 50 volts (A.C.). Above this limit, the ratio of moment produced by the element to applied voltage decreases, and at very high applied voltage, the element depolarizes. Power considerations are important in the space applications. The piezoelectric elements require relatively large voltages up to 50 volts, but the current they draw is on the order of micro-amps. Thus, the power consumed by the vibration control system is not expected to be prohibitive. For example, if each transducer requires 50 volts and draws 10 micro-amps, the power consumed by 2000 sensor/driver pairs is 1 Watt.

BACKGROUND

A perfectly bonded piezoelectric may be modeled as being pinned at either end of the element. A force proportional to the applied voltage is exerted through these pins to the structure. Expressed in terms of modal

parameters, the equation of motion is of the form:

$$M*\ddot{q} + C*\dot{q} + (K+K_p)*q = F \qquad (1)$$

where:
M= the modal mass
C= the modal damping inherent in the structure
K= the modal stiffness
K_p= the stiffness of the piezoelectric element (typically small in comparison to the modal stiffness of the structural member)
q= the modal displacement
F= the force exerted by the piezoelectric element transformed into modal coordinates.

The expression for the force exerted is derived by Crawley and deLuis [1], and is based on the force term shown for a monomorph motor application in the Program Module for Piezo Design Aid [2]. The force has the form:

$$F = K_1*d_{31}*V*\left(\frac{d\phi}{dx}\Big|_{a+h} - \frac{d\phi}{dx}\Big|_{a-h}\right) \qquad (2)$$

where:
K_1= a constant incorporating physical parameters of the structural member and of the piezoelectric element, such as size and modulus of elasticity.
d_{31}= the piezoelectric constant.
V= the voltage applied across the piezoelectric element.

$\frac{d\phi}{dx}\Big|_{a+h} - \frac{d\phi}{dx}\Big|_{a-h}$ = the spatial derivative of the mode shape at the end points of the piezoelectric element.

The piezoelectric element used is lead zirconate barium titanate ceramic compound, designated G-1195. This material has a relatively high piezoelectric constant. It also has an elastic modulus nearly equal to that of aluminum. All beams in this project are fabricated from aluminum. The equality of moduli of elasticity between the piezoelectric element and the structure to which it is bonded is a recommended design parameter. Each piezoelectric element used is 1.00 inch (25.4 mm) long X 0.25 inch (6.3 mm) wide X 0.01 inch (0.25 mm) thick.

When the piezoelectric is used as a transducer, it produces a voltage which is proportional to the rate of stress at either end of the element. The proportionality is based on the strain gage relationship in the Piezo Design manual [2]; and has the form:

$$V_{out} = K_2*g_{31}*c*\frac{d}{dt}\left(\frac{d^2y}{dx^2}\Big|_{a-h} - \frac{d^2y}{dx^2}\Big|_{a+h}\right) \qquad (3)$$

where:
K_2= a constant incorporating the material properties of the piezoelectric element.
g_{31}= the voltage coefficient.
c= distance to the neutral axis.

It has been assumed that the stress in the piezoelectric element is the same as the stress at the surface of the structural element, and the piezoelectric has no effect on the actual stress at the member surface. While this assumption is not strictly correct, it is expected that the relationship gives an approximation of the voltage produced by the piezoelectric.

The relationships derived above give the expected values both of the voltage produced by the piezoelectric element (as transducer) and of the force produced by the element (as driver).

RESULTS

A series of experiments was performed in order to establish the effect of bonding technique on the performance characteristics of the piezoelectric elements. These tests used a cantilevered beam with the sensor/driver pair mounted near the root. The beam is 6061-T6 aluminum, 0.75 inch (19.1 mm) wide X 0.125 inch (3.2 mm) thick, ranging in length from 15.13 inch (384 mm) to 17.84 inch (453 mm). The bonding parameters varied were the adhesive used to affix the piezoelectric to the beam and the type of electrical lead.

Mounting of the piezoelectric can be achieved with most bonding techniques [2]. However the voltage produced by the sensor or the moment exerted by the driver is affected by the hardness of the bond. Adhesives used for bonding are a 910-type pressure sensitive contact adhesive and a 24-hour cure two part epoxy. The piezoelectric element has a positive and a negative pole. The surface is coated with a nickel electrode to which leads are attached to get voltage out of or into the element. Small tabs of aluminum foil were used in the initial tests. However, a suitable means to affix the wires leading from the tabs could not be found. One mil thick brass tabs were then fabricated. The lead wires could be soldered to the brass tabs. A technique was perfected to solder the lead wires directly to the piezoelectric electrodes. This necessitates drilling a small groove in the beam in order to make room for the solder joint.

It is necessary to isolate the piezoelectric from the beam in order to eliminate ground loop effects. Thus, each piezoelectric is bonded in the following manner. An insulating interlayer is bonded to the beam. The leads are attached to the electrodes of the piezoelectric, and the piezoelectric is bonded to the insulating interlayer. For four of the five beams for which results are reported, the interlayer is Mylar tape. Beam number 4 uses an interlayer of tissue that is soaked with the 910 adhesive.

In the first series of tests, the piezoelectric was driven at steady state at the first resonant frequency of the beam to which it was bonded. The steady state deflection of the beam tip was recorded as a function of drive voltage. Figure 1 shows the test setup. The results of a typical test are shown in figure 2. The tip deflection increases linearly with applied voltage up to approximately 50 volts. Beyond this voltage, the rate of increase of tip deflection flattens. The maximum voltage from the amplifier is 59 volts. The measured results are compared to expected results calculated as follows.

The first bending mode of a cantilever beam can be represented by the mode shape function:

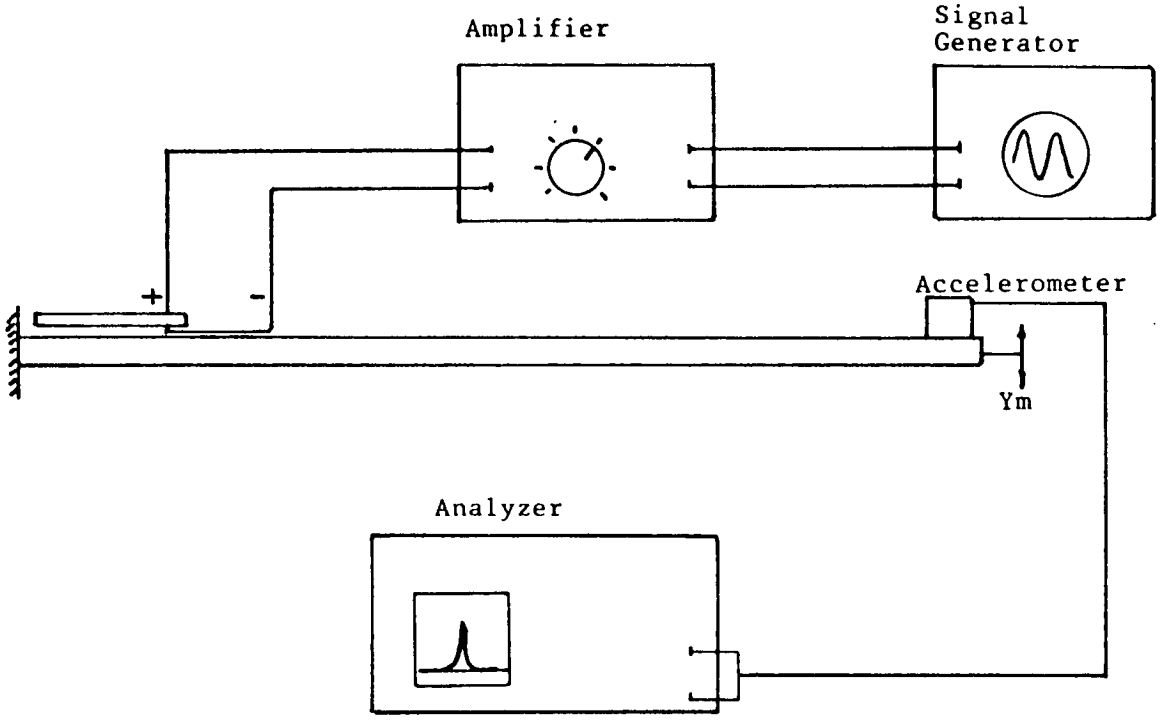


Figure 1. Test Setup to Determine Steady State Tip Deflection for Cantilever Beam Driven by Piezolectric

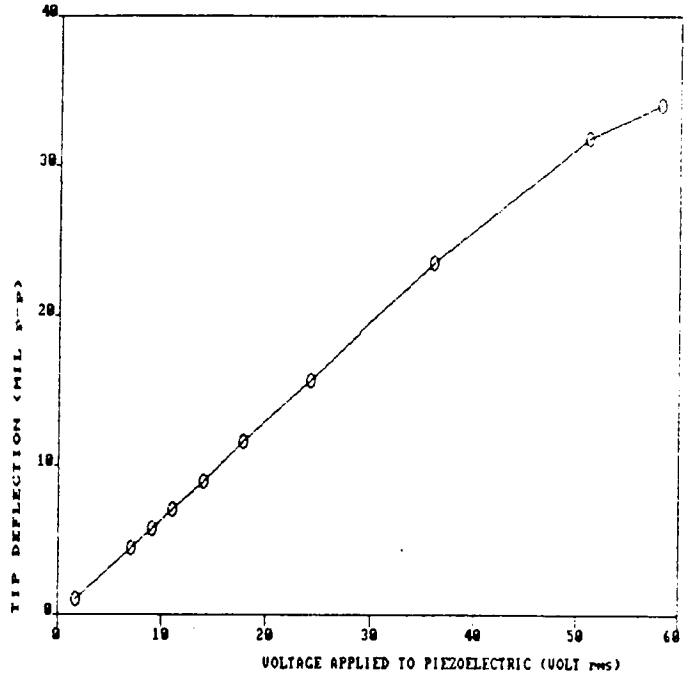


Figure 2. Steady State Tip Deflection of Excited Cantilever Beam

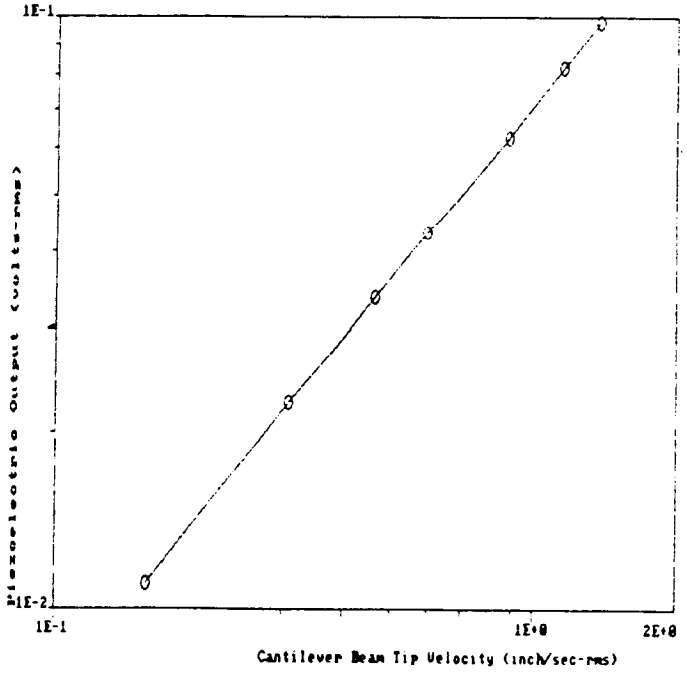


Figure 3. Steady State Voltage Signal from Piezoelectric on Cantilever Beam

Table I. Steady State Response at Cantilever Beam Tip For Excitation at First Resonance.

Aluminum 6061-T6, W = 0.75 in, t = 0.125 in.

BEAM	LENGTH (in)	ELECTRODE	BOND	K (lb/in)	ζ	F_n (Hz)	A_{exp} $X10^4$ (IN/V)	A_{meas} $X10^4$ (IN/V)
1	16.844	Aluminum Foil	24-hour Epoxy	3.065	0.0148	14.0	2.92	2.74
2	17.675	Brass	24-hour Epoxy	2.675	.0167	12.75	3.945	1.616
3	15.250	Soldered	Top 24-hour Epoxy Bottom 910	4.130	.00885	16.20	4.864	3.680
4	15.125	Soldered - to beam-top + to beam-bottom	910 adhesive Tissue inter-layer	4.233	.00662	17.00	6.350	12.00
5	15.250	Soldered - to beam-top + to beam-bottom	910 adhesive	4.130	.00716	16.00	6.01	6.370-6.600

$$y(x,t) = q(t) * \phi(x) \quad (4)$$

where:

q(t)= displacement of the tip of the beam.

and:

$$\phi(x) = \frac{3}{2}\left(\frac{X}{L}\right)^2 - \frac{1}{2}\left(\frac{x}{L}\right)^3 \quad (5)$$

L= length of the beam.

Equation 5 is derived assuming that the dynamic mode shape is the same as the static mode shape.

When the forcing function in equation 1 is a sinusoid of frequency Ω and magnitude F_o, the temporal response q (t) of the beam tip is also a sinusoid at the same frequency. When the excitation frequency matches the resonant frequency of the beam, the magnitude of the tip deflection is:

$$Y_m = F_o/(2 * \zeta * (K+K_p)) \quad (6)$$

where:

ζ = the damping coefficient of the beam in the first mode. For a lightly damped system, this is the maximum tip deflection.

The expected modal force is calculated from equation 2 using the physical parameters of the beam and piezoelectric, and the assumed mode shape, equation 5. Note that the modal force is directly proportional to the applied voltage. The modal damping is measured from free vibration test. The modal stiffness is estimated, using the assumed mode shape, from:

$$K=3*E*I/L^3$$

The stiffness of the piezoelectric is evaluated using relationships found in reference [1] and is found to be K =0.091 lb/in. This value is on the order of 1/30 of the range of beam stiffnesses.

These values are combined in equation 10 to estimated the tip deflection as a function of the applied voltage.

The expected and measured values of the tip deflection for the five beams are summarized in table I. The measured tip deflection is within +/- 10% of the expected value for beams 1 (aluminum foil leads) and 5 (soldered leads, 910 adhesive). The brass leads, beam 2, do not permit sufficient voltage to pass the electrodes. Beam 3 has soldered leads. The sensor piezoelectric is affixed with 24-hour epoxy, while the driver is affixed with the 910 adhesive. Except for the difference in adhesive, the mounting technique is essentially the same as for beam 5. It is not clear why beam 3 does not show the expected tip deflection. Beam 4 shows nearly twice the expected tip deflection. This is due to the elimination of Mylar interlayer, which greatly improves the hardness of the bonding.

In the next test, the voltage produced by the piezoelectric was monitored while its companion was driven at steady state. It is expected from equation 3 that the voltage output from the piezoelectric is proportional to the rate of strain. The strain is the second spatial derivative of the beam displacement, which is the tip displacement times the mode shape. The mode shape is constant in time in the first mode.

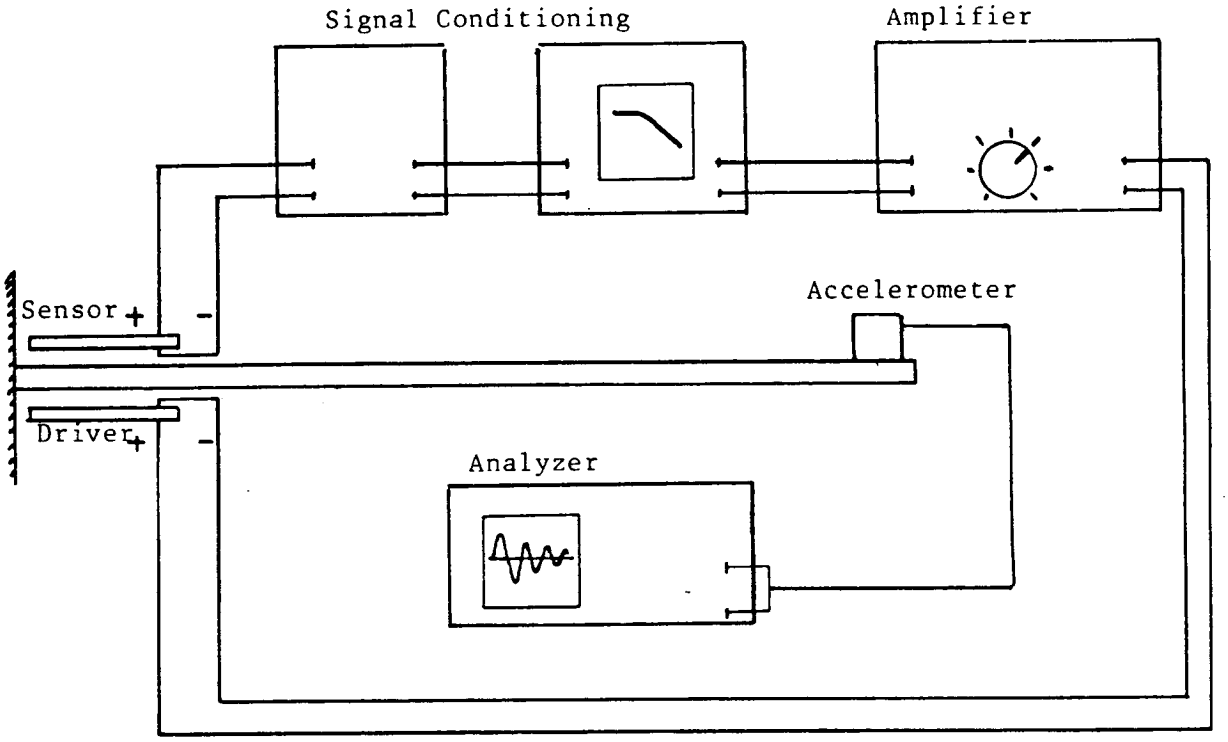


Figure 4. Feedback Control System for Cantilever Beam

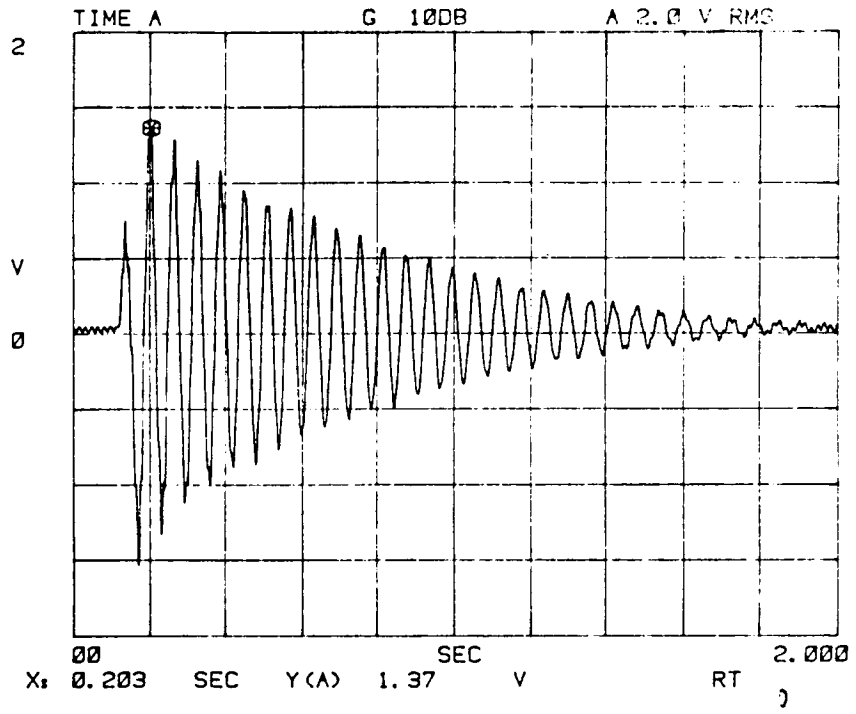


Figure 5. Step Input Response at Cantilever Beam Tip
-Controller Off

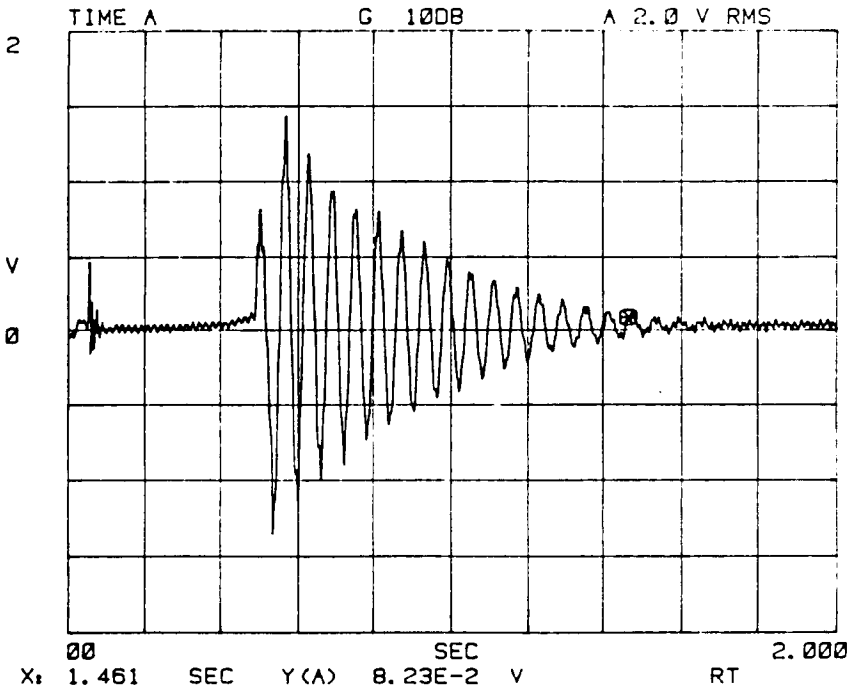


Figure 6. Step Input Response at Cantilever Beam Tip
- Controller On

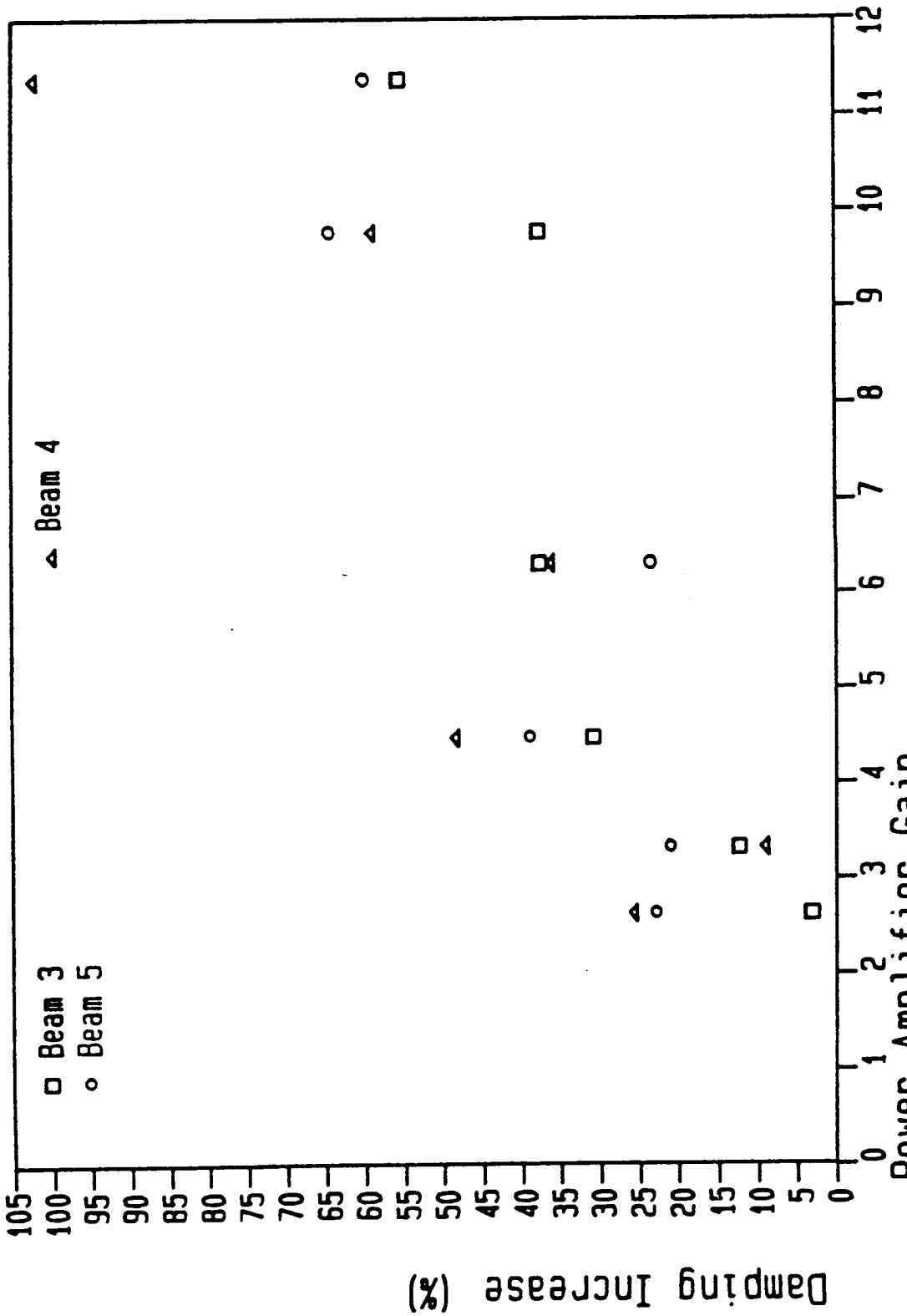


Figure 7. Effect of Bonding Technique on Damping Efficiency

Thus, the rate of strain is proportional to the tip velocity. Figure 3 shows the relationship between voltage output from the piezoelectric and tip velocity. This relationship is linear with a slope of 0.0719 volts/inch/sec. The expected value is obtained from equation 3, where the second derivative of the beam deflection with respect to x is evaluated using the assumed mode shape. The expected slope is 0.0980 volts/inch/sec. The measured value is 25% less than the calculated. This is felt to be due to the fact that the derived value is based on a configuration in which two piezoelectric elements are used to form a strain gage. Thus the model is somewhat different from the experiment. The measured and expected results are of the same order of magnitude. Better correlation is expected when the model is improved.

In the final series of preliminary tests, the piezoelectrics are configured in a sensor/driven pair, as shown in figure 4. The voltage output from the sensor is passed through signal conditioning which shifts the phase by 180°, filters out 60 Hz noise, and amplifies the voltage. The conditioned voltage is passed to the driver. The beam tip is displaced by approximately 0.25 inch, and released from rest. Figure 5 shows a typical free response of the beam from an accelerometer mounted at the tip. The free response is obtained with the feedback control system turned off. When the feedback control system is activated, the response to an initial displacement is as shown in figure 6. It is seen that the feedback control system produces a significant increase in damping.

Beams 3, 4, and 5 were evaluated to determine the increase in damping efficiency as a function of feedback gain factor. These three beams were chosen because beam 2, with the brass leads had already been eliminated, based on the previous series of tests. Beam 1, with the aluminum leads was evaluated, but produced no noticeable increase in damping.

Figure 7 shows the increase in damping with amplifier gain, for factors up to approximately 12. The gain recorded is for the amplifier only. The signal conditioner has a gain of 16.3 at 16.0 Hz. Thus, the total amplification of the signal out of the sensor piezoelectric is as much as 192. The figure shows that the increase in damping generally follows the signal amplification. An anomalous dip in the general trend at the gain factor of 6 is felt to be due to the power amplifier. Beam 4, using the tissue interlayer, appears to have the greatest damping efficiency. However, it was found that at high amplification, around 10, the beam became unstable, with a vibration mode at 104 hertz being self-excited.

Equation 1 is used to evaluate the results of this feedback test. The modal force produced by the drive piezoelectric is proportional to the voltage applied to it. The applied voltage is the amplified voltage produced by the sensor piezoelectric. The sensor voltage is proportional to the time derivative of the strain, which is the modal velocity times the second spatial derivative of the mode shape. Thus, the feedback control signal is proportional to the modal velocity.

The beam is excited into its first resonant mode by an initial displacement of the tip. The equations of motion is:

$$M * \ddot{q} + C * \dot{q} + (K+K_p) * q = -T_1 * T_2 * G * \dot{q} \qquad (7)$$

where:

T_1 = driver proportionality term

TABLE II. DAMPING EFFICIENCY OF FEEDBACK CONTROL SYSTEM - BEAM 5

POWER AMPLIFIER GAIN	OVERALL SYSTEM	EXPECTED		MEASURED
		ζ EFF	% INCREASE	% INCREASE
2.58	42.05	.00133	19	23
3.30	53.79	.00170	24	22
4.45	72.54	.00229	32	39
6.31	102.9	.00325	45	24
9.75	158.9	.00503	70	65
11.28	183.9	.00582	81	60

T_2 = sensor proportionality term
G^2 = amplified gain

The excitation term is transferred to the left hand side of the equation. Dividing by the modal mass, equation 7 becomes:

$$\ddot{q} + 2 * w_n^2 * (\zeta + \zeta_{eff}) * \dot{q} = w_n^2 * q = 0 \qquad (8)$$

Thus, the feedback signal adds an effective modal damping to the system.

Table II is a summary of the damping increase as a function of power amplifier gain. The signal conditioner circuitry provides a fixed gain of 16.3 at the free vibration frequency of beam 5. The expected values of effective damping are calculated from equation 10. These are used with the inherent damping in beam 5, shown in table I to be 0.00716, to calculate the expected damping increase. The measured values are taken from figure 7. The expected and measured values generally agree to within +/- 7%. The wide divergence at power amplifier gain of 6.3 coincides with a general dip in the curves in figure 7. The measured damping efficiency is 20% less than the expected at the gain of 11.3. This is felt to arise from the fact that, at such high gain, the drive voltage exceeds 50 volts, and the piezoelectric efficiency degrades.

Thus, a simplified algorithm has been developed to estimate the transient response of the beam as a function of feedback amplification. It is expected that this algorithm can be modified for the free-free beam case in order to verify the experimentally obtained results and to aid in the parameter evaluation phase of the project.

TESTS ON FREE-FREE BEAM

The purpose of the preliminary work is to establish the bonding techniques and to aid in the design of feedback circuitry for the free-free beam. Based on the preliminary work, it was decided to mount the piezoelectrics to the beam with the 910 adhesive using a Mylar tape interlayer. The electric leads are soldered to the electrodes of the piezoelectric. The free-free beam experiment is shown schematically in figure 8. The beam is 96 inches long and is suspended vertically from one end. The pendulum length is approximately 25 feet.

The purpose of the long pendulum is to simulate a free end condition with minimum shear and moment. The pendulum frequency is expected to be below 1.0 hertz. This frequency is below the first flexural model of the beam, which is calculated to be at 2.77 hertz. Eight sensor/driver pairs of piezoelectrics are mounted to the beam at intervals of 12 inches. The sensor signal for each piezoelectric pair is amplified and shifted in phase by 180° before being reintroduced to the driver. The signal conditioning also includes a low pass filter to reduce 60 hertz noise. The total gain of each amplifier is 230. In order to ensure that the piezoelectric is not overdriven, the power supply to each amplifier is limited to +/= 70 volts, or approximately 50 volts rms.

A special striker was fabricated to ensure that the beam is excited with the same initial energy each time. Figure 9 shows the free vibration of the beam with the feedback system off. The beam was struck in the center. It was found that the accelerometer signal response was attenuated at the first resonance of 2.77 hertz, and did not give a true

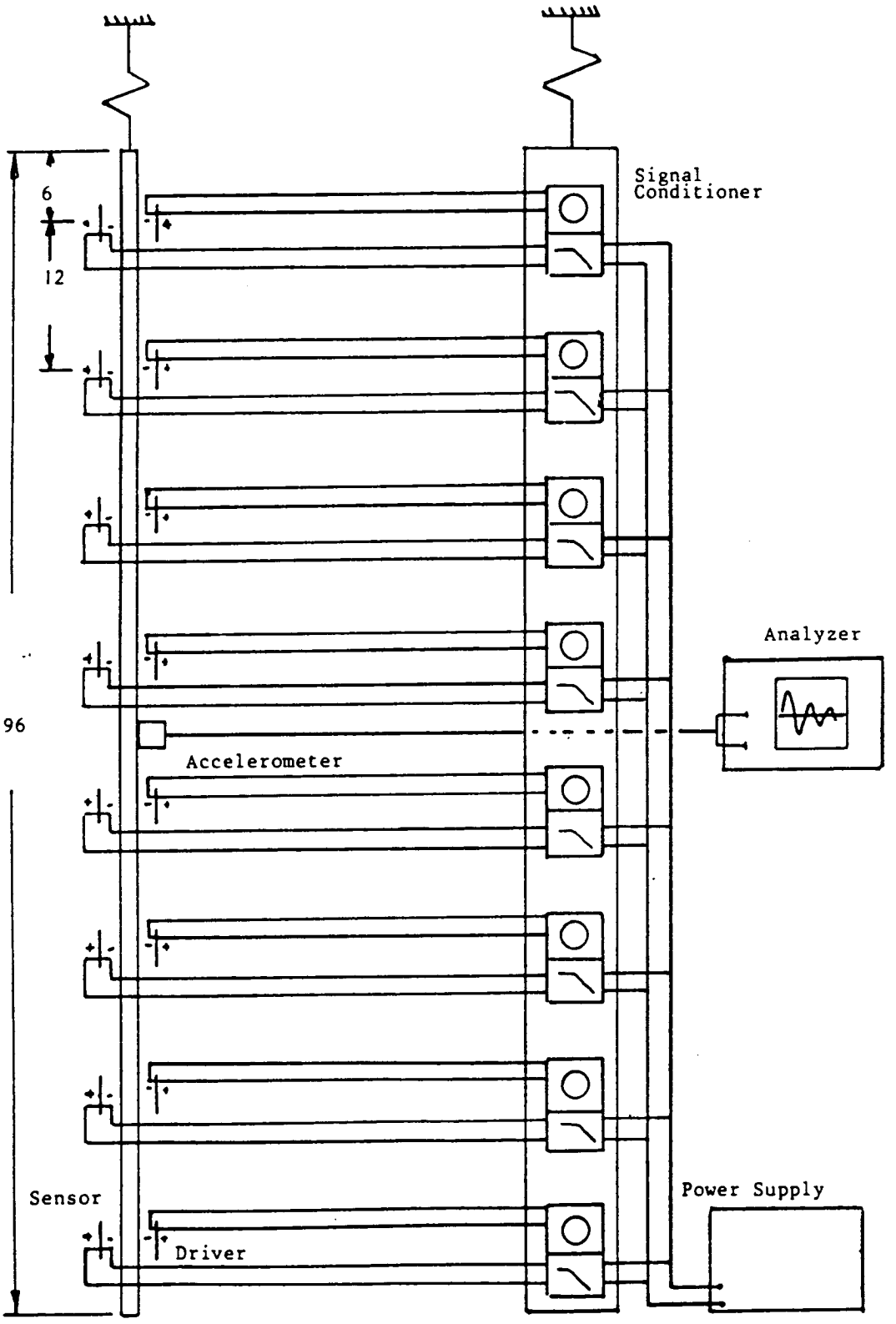


Figure 8. Test Setup for Free-Free Beam Controller

representation of the beam vibration in the first mode. Thus, the signal from a piezoelectric near the center of the beam was used for measurement of vibration. The result obtained with the control system on is shown in figure 10. It can be seen that the vibrations damp out much more quickly than they do with no feedback. The peak displacement of the beam is approximately 1/2 the peak displacement with no control.

The excitation method excites more than one vibration mode in the beam. Figure 11 shows the spectrum of beam oscillation when the beam is excited in the center. The vibration peaks are harmonically related. It is expected that the peaks would occur at odd harmonics of the fundamental (2.77 hertz, 14.95 hertz, 36.96 hertz, and 68.74 hertz) . While this result is generally found, peaks also occur at the odd harmonics of approximately 17.5 hertz (52.5 hertz, and 87.5 hertz). These vibration components arise from transverse vibration in the cable of the pendulum supporting the beam. These vibration components happen to coincide with the beam's flexural components. Since the transverse vibration of the cable is not affected by the piezoelectric control system, this vibration energy adds to the flexural energy in the beam, and degrades the damping efficiency of the piezoelectrics. It is expected that modification of the beam suspension will eliminate the transverse response, and the damping efficient of the feedback control system will be more clearly demonstrated.

Since multiple modes of the beam are excited, evaluation of modal damping is not possible. However, an approximation of the overall energy reduction is made. It is assumed that the area under the response curve is related to the energy. The area is approximated by faring a curve through the peaks of the response. The area under the curve in figure 10 is 47 percent of the area under the curve in figure 9. This indicates that the energy input by the piezoelectric controllers reduces the overall response by a factor of 2.13.

Striking the beam at the end excites both the even and odd harmonics. The results obtained are similar to the case in which the beam was excited at the center. As was the case with excitation in the center of the beam, it is expected that the damping efficient of the feedback control system will be more clearly demonstrated when the transverse cable vibration is eliminated.

CONCLUSIONS

Practical considerations, including techniques for mounting the piezoelectric elements and for attaching electrical leads have been addressed. It is found that bonding the element to the beam using a hard adhesive such a 910 type is suitable because stresses at the beam surface are transferred through a hard bonding layer most efficiently . Mylar tape is used as an insulator to ensure that the element is not grounded to the beam. Soldering the electric leads to the piezoelectric element ensures efficient transfer of the high voltages at low currents either generated by the piezoelectric or used to drive it.

An analytical model of the feedback control system has been developed. This model estimates the voltage generated by the piezoelectric sensor as a function of the dynamic stress at the sensor location, and the force exerted by the driver piezoelectric as a function of signal gain. Both

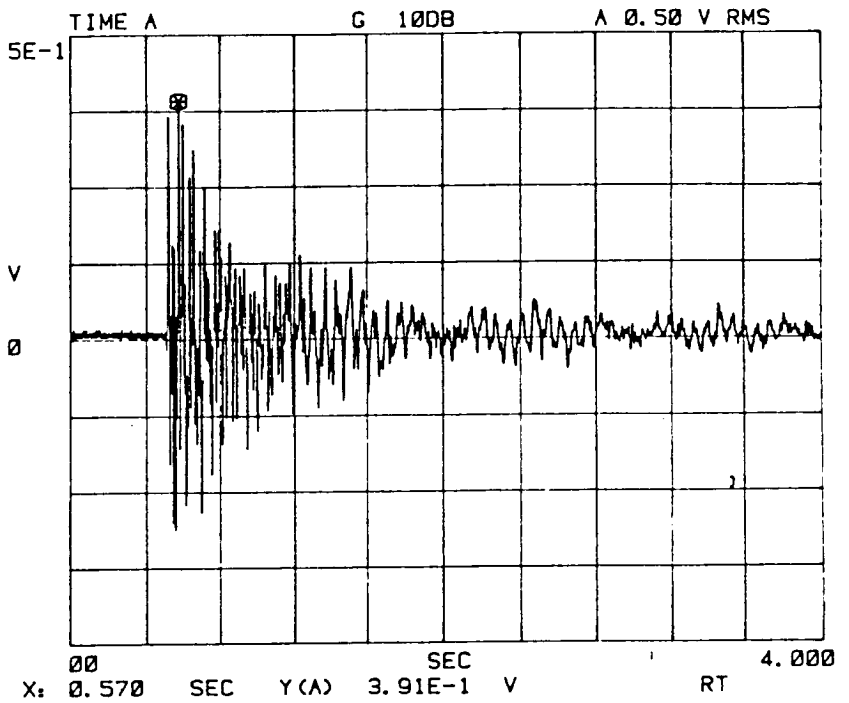


Figure 9. Free-Free Beam Excited at Center - Control System Off

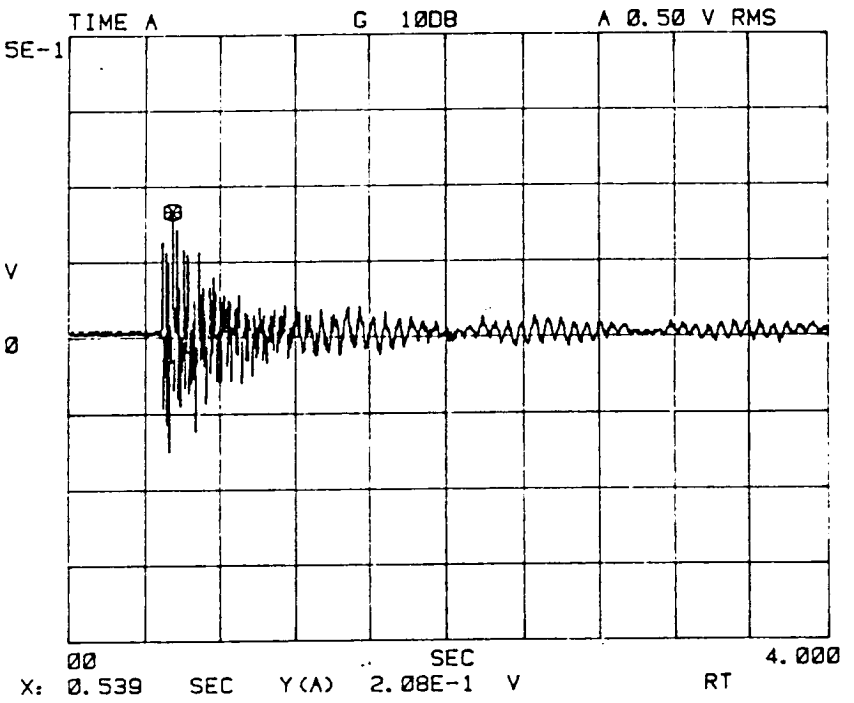


Figure 10. Free-Free Beam Excited at Center - Control System On

C-3

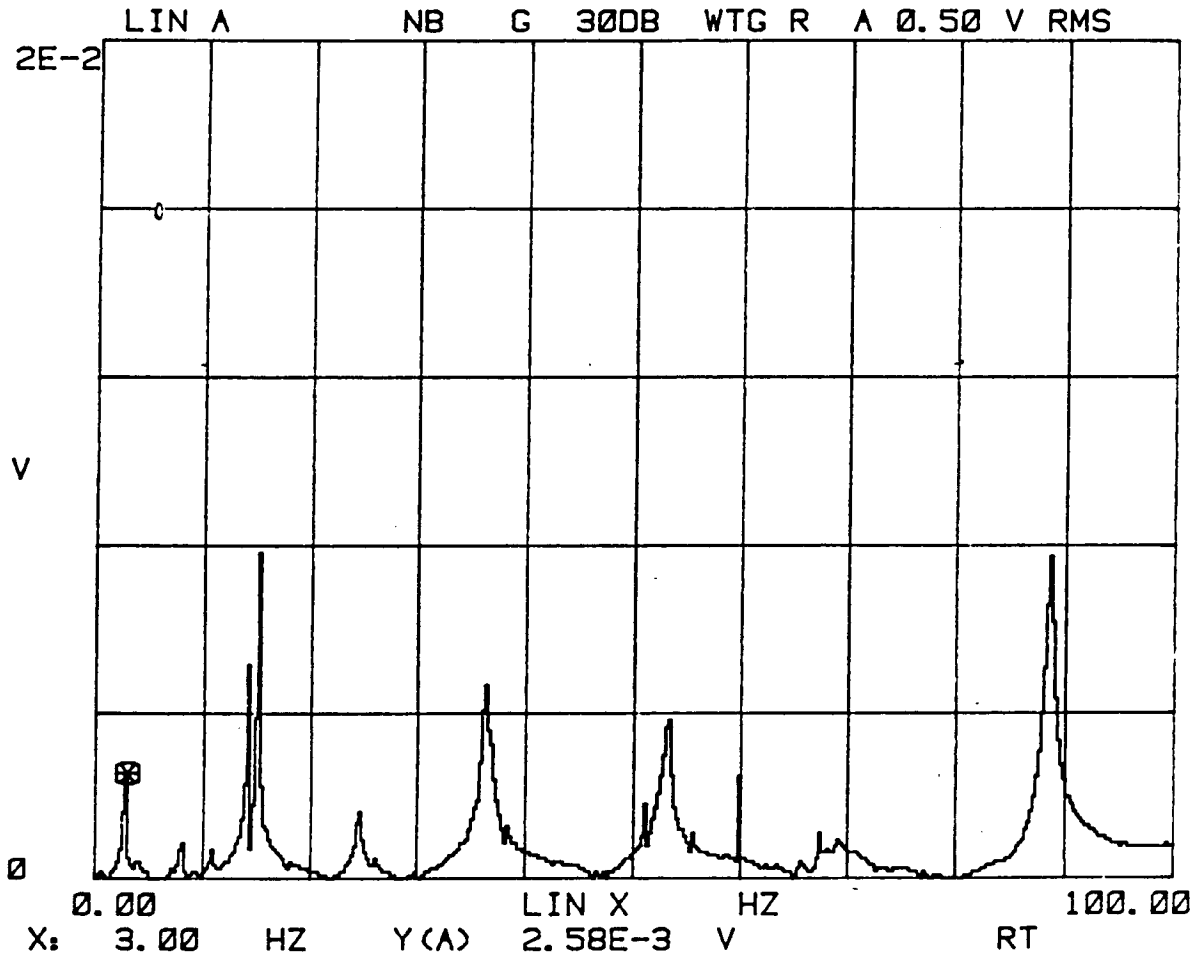


Figure 11. Spectrum of Free-Free Beam Response

terms are functions of the mode shape of beam vibration. The feedback control system is expressed in terms of a modal damping factor. The model has been compared to measured results for a cantilever beam excited to vibrate in its first natural mode. The estimated increase in first modal damping factor compares favorable wit the measured results.

A free-free beam configuration has been simulated and instrumented with a distribution of piezoelectric sensor/driver pairs. The purpose of the distribution is to test the vibration control efficient of the piezoelectrics when they are not optimally located at points of high stress in the beam. The feedback control system is shown to reduce the response to impulsive excitation by a factor of approximately 2.

ACKNOWLEDGMENTS

The author wishes to acknowledge the personnel of the Loads and Dynamics and of the Vibration and Acoustics Sections for their help and encouragement in the performance of this project. The aid in fabrication of the test setup provided by technical personnel at the Vibration and Acoustics facility is also gratefully acknowledged. Finally, the author expresses his appreciation to NASA and ASEE for providing the opportunity to work on this project.

REFERENCES

1. Crawley, E.F., and deLuis, J., "Experimental Verification of Distributed Piezoelectric Actuators for Use in Precision Space Structures", AIAA paper number 86-0878, 1986.

2. "Piezo Design Aid, volume 1.", Piezoelectric Products, Incorporated 212 Durham Avenue, Metuchen, NJ, 08840, 1985

PREDICTION OF PHYSICAL WORKLOAD IN REDUCED GRAVITY ENVIRONMENTS

Final Report

NASA/ASEE Summer Faculty Fellowship Program -- 1987

Johnson Space Center

Prepared by: Joseph H. Goldberg, Ph.D.

Academic Rank: Assistant Professor

University & Department: The Pennsylvania State University
Department of Industrial Engineering
207 Hammond Bldg.
University Park, Pennsylvania 16802

NASA/JSC

Directorate: Engineering

Division: Advanced Programs Office

Branch: Systems Definition

JSC Colleague: John W. Alred, Ph.D.

Date: August 14, 1987

Contract Number: NGT 44-001-800

ABSTRACT

This paper describes the background, development, and application of a methodology to predict human energy expenditure and physical workload in low gravity environments, such as a Lunar or Martian base. Based on a validated model to predict energy expenditures in Earth-based industrial jobs, the model relies on an elemental analysis of the proposed job. Because the job itself need not physically exist, many alternative job designs may be compared in their physical workload. The feasibility of using the model for prediction of low gravity work was evaluated by lowering body and load weights, while maintaining basal energy expenditure. Comparison of model results was made both with simulated low gravity energy expenditure studies and with reported Apollo 14 Lunar EVA expenditures. Prediction accuracy was very good for walking and for cart pulling on slopes less than 15°, but the model underpredicted the most difficult work conditions. This model was applied to example core sampling and facility construction jobs, as presently conceptualized for a Lunar or Martian base. Resultant energy expenditures and suggested work-rest cycles were well within the range of moderate work difficulty. Future model development requirements were also discussed.

1. INTRODUCTION

1.1. Need for Work Prediction Tools

The National Commission on Space (1986) has recently strongly recommended that the U.S. proceed in a research program to "support human settlements beyond Earth orbit, from the highlands of the Moon to the plains of Mars (p. 193)". If working in low gravity space environments is to become commonplace, the physical workload imposed by everyday tasks must be well understood. Based on extrapolations from previously developed models, we now have an opportunity to develop low gravity job design tools to insure that workloads imposed by tasks will not be beyond an average worker's capability. Although many aspects of work are becoming more cognitive and less physical in nature, the initial population of the Moon and other planets will require a great deal of physical work. Prediction of the physical workload of low gravity physical jobs will allow an estimation of the amount of rest required during a given job shift. In turn, the required amount of EVA surface time can be estimated. This paper presents a technique, extrapolated from Earth-based models of human energy expenditure, for predicting the physical workload imposed by a job in low gravity environments. Although many of the models likely will require significant modification for lower gravity work prediction, the elemental job analysis methodology is quite applicable.

1.2. Work Efficiency, Job Design, and Energy Expenditure

Humans are relatively inefficient when compared to working machines. Under the most optimal conditions, we can convert about 30% of input food energy to useful work,

with the remaining 70% is wasted as heat (Sanders and McCormick, 1987). The design of a working task, nutrition, and other personal health factors dictate whether work is performed at a low or high efficiency. For example, shoveling in a stooped posture has an efficiency of about 3%, whereas shoveling in a more erect posture has about 6% energy efficiency (Grandjean, 1981). Webb (1973) reported that respiration is relatively inefficient at less than 5%, common tasks are 10-20% efficient, whereas bicycling and walking on an inclined treadmill may be upwards of 35% efficient in converting input energy to useful work. Careful design of jobs can insure that an individual is operating at the highest possible efficiency level, by lowering energy expenditure levels as much as possible.

Measurement of heat production provides a basis for estimating the efficiency of work. The rate of energy expenditure, in Kilocalories/minute (Kcal/min) can, in fact, provide a convenient measure of the difficulty of work. A resting individual expends about 1.5-2.5 Kcal/min, or about 1440-2400 Kcal per 16 hour waking day, just to perform the body's vital functions.[1] This basal expenditure is always present. On top of one's basal energy expenditure, additional energy is required to perform useful work, required when the body's limbs are moved through a distance. These additional task-based expenditures can exceed 10 Kcal/min for grueling work. The physical demands of a job or task may be defined through a consensus understanding of energy expenditure limits by the body. A well-accepted gradation of work demand was published by the American Industrial Hygiene Association (1971), and is shown in Table 1. The stated expenditures, for a healthy adult male, include both basal and task-based energy expenditures (E's), and vary from resting (1.5 Kcal/min) to Unduly Heavy Work (>12.5 Kcal/min). These E's may be physiologically measured, as described below. The example tasks, from Webb (1973), allow a more subjective understanding of the work difficulty. Writing while sitting at a table is essentially resting or basal E, whereas walking in loose snow can be unduly heavy work.

1.3. Objectives

This study has the objective of proving the feasibility of predicting physical workload in low gravity tasks by extrapolating from an Earth-based energy expenditure prediction model. After initial model development, an error analysis will add corrections to the model. As an example of model application, a sample task presently being designed for a Lunar or Martian base will then be subjected to workload and work-rest cycle analysis.

---

[1]Nutritional intake must compensate for this energy expenditure. A common misconception is that calories, as opposed to kilocalories, are the basic common unit of food energy.

TABLE 1. WORK GRADE AND ENERGY EXPENDITURES, ADAPTED FROM AIHA (1971)

WORK GRADE	$\dot{E}$ (KCAL/MIN)	ENERGY/8 HR (KCAL/8 HR)	HEART RATE (BEATS/MIN)†	SAMPLE TASKS
Rest (sitting)	1.5	<720	60-70	Writing
Very Light Work	1.6-2.5	768-1200	65-75	Riding in Car/Typing
Light Work	2.5-5.0	1200-2400	75-100	Slow Walk/Lecturing
Moderate Work	5.0-7.5	2400-3600	100-125	Crawling/Tennis
Heavy Work	7.5-10.0	3600-4800	125-150	Chopping Wood/Stair Climb
Very Heavy Work	10.0-12.5	480-6000	150-180	Basketball/Cycling
Unduly Heavy Work	>12.5	>6000	>180	Wrestling/Walking in snow

†Note: These are typically observed heart rates; actual rates will vary highly
*Source: Webb (1973)

2. CURRENT JOB DESIGN TOOLS

2.1. Measurement of Energy Expenditure

The energy expended by a working individual may be inferred by measuring rate of O_2 consumption and/or CO_2 production while performing a task. The volume of expired air is measured during a specified time period, and the percentage of O_2 and/or CO_2 in this sample is taken. The difference in O_2/CO_2 percentage between inspired and expired air indicates percentage utilization/production. When multiplied by volume of expired air and divided by time, the rate of O_2 consumption/CO_2 production in volume/time is obtained. For a healthy, working adult, with normal nutrition and metabolism, about 4.9 Kcal of energy is generated for each liter of O_2 consumed/CO_2 produced (Ästrand and Rodahl, 1977). Other methods of physiological determination of energy expenditure also exist, such as direct measurement of heat generation, or inference from heart rate. Lunar EVA expenditures were generally determined from oxygen consumption, after correcting for suit leakage and other factors (see Johnston, et al., 1975).

2.2. Prediction of Energy Expenditure

Energy expenditures in a task may be predicted via a methodology introduced by Garg (1976), and Garg, Chaffin, and Herrin (1978). Using extensive measurement of O_2 utilization by six young, healthy subjects performing controlled manual labor, these investigators derived equations expressing energy expenditure associated with specific job elements. Their regression equations express energy expended as a function of such personal parameters as body weight, weight of load, walking velocity, etc. The form of this prediction model is:

$$\bar{\dot{E}}_{job} = \frac{\sum_{i=1}^{n_{pos.}} \dot{E}_{posture_i} \times t_i + \sum_{i=1}^{n_{task}} \Delta E_{task_i}}{T} \qquad (1)$$

Where:

$\bar{\dot{E}}_{job}$ = Ave. job E (Kcal/min)

$\dot{E}_{posture_i}$ = E due to maintenance of ith posture (Kcal/min)

ΔE_{task_i} = E due to ith task during steady state work (Kcal)

$n_{pos.}$ = Number of postures employed in job

n_{task} = Number of tasks in job

t_i = Time duration of ith posture (Min)

T = Time duration of job (min)

Equation 1 summates basal energy expenditure (left side of numerator) with energy expenditure due to specific tasks (right side of numerator). In the model, there are three basal postures: sitting, standing, and standing bent, as defined in Table 2. Body weight is simply multiplied by a constant for each of these postures, to achieve basal energy expenditures in the area of 1.5-2.5 Kcal/min. Each of the three basal E's are multiplied by the fraction of time spent in that posture (t_i/T) and summed to achieve the final basal E. Task energy expenditures are defined via regression equations for each job element. These are shown in Table 1, beneath the basal energy equations. Note that the task ΔE equations shown represent only a selected fraction of the equations in the model. Walking and Carrying yield Kcal spent in a duration of time, t. The Lifting and Lowering equations shown here describe energy expended per lift/lower, while keeping the legs essentially straight. The Pushing/Pulling equation describes energy expended during a movement of horizontal distance, X. Comparisons between alternative task arrangements made be made by altering the various input parameters, and noting overall impact on average task energy expenditure. In general, this model is extremely accurate and sensitive in predicting the energy expenditure of Earth-based industrial jobs (see Garg, Chaffing, and Herrin, 1978).

2.3. Work-Rest Allocation

Once the energy expenditure of a task has been defined, the next consideration is how much rest to include as part of a shift, to avoid excessive fatigue. An accepted technique in the industrial community relies on the concept of physical work capacity (PWC), the maximum work effort that an individual (or population) can exert over a shift, without any rest time, and without incurring any short or long term injury. As a rule of thumb, the energy expenditure rate at PWC (E_{pwc}) should be 1/3 one's maximum aerobic capacity for 8 hour shifts, and 1/2 one's maximum aerobic capacity for 4 hour shifts

(see Bink, 1962). For shorter shifts, higher E_{pwc}'s may usually be sustained without incurring any injury. By definition: $T_{job}\ E_{pwc} = T_{rest}\ E_{rest} + T_{work}\ E_{work}$, and $T_{job} = T_{rest} + T_{work}$, so after solving for rest time,

$$T_{rest} = T_{job} [E_{work} - E_{pwc}] / [E_{work} - E_{rest}] \qquad (2)$$

Where: T_{rest} = rest time (min) E_{work} = working E (Kcal/min)
T_{job} = total shift time (min) E_{rest}= basal E (Kcal/min)
T_{work} =work time (min) E_{pwc} = pwc E (Kcal/min)

TABLE 2. SELECTED ENERGY PREDICTION MODEL ELEMENTS, FROM GARG (1978)

BASAL ENERGY EXPENDITURES (KCAL/MIN)
Sitting: E = .023 {BW} Standing: E = .024 {BW} Standing Bent: E = .028 {BW}
WALKING AND CARRYING TASK ENERGY EXPENDITURES (KCAL)
Walking: $\Delta E = .01 \{ 51 + 2.54 [BW \times V^2] + .379 [BW \times G \times V] \} t$
Carrying (at Waist):
$\Delta E = .01 \{ 68 + 2.54 [BW \times V^2] + 4.08 [L \times V^2] + 4.62 [L] + .379 [L+BW] [G \times V] \} t$
LIFTING AND LOWERING ENERGY EXPENDITURES (KCAL/LIFT, LOWER)
Stoop Lift: $\Delta E = .01\{.325 [BW] [.81-H1] + [1.41 (L) + .76 (S \times L)] [H2-H1]\}$
Stoop Lower: $\Delta E = .01\{.268 [BW] [.81-H1] + .675 (L) (H2-H1) + .522 (S)(.81-H1)]\}$
PUSHING/PULLING AT .8 M HEIGHT (KCAL/PUSH)
Pushing/Pulling: $\Delta E = .01 [X] \{.112 [BW] + 1.15 [F] + .505 [S \times F] \}$

Where:	BW = Body Weight (Kg)	L = Weight of Load (Kg)
	F= Ave. Pushing/Pulling Force (Kg)	S = Gender (Males = 1, Females = 0)
	G = Grade of Surface (%)	V = Speed of Walking (M/sec)
	H1, H2 = Vertical Height from Floor (M)	X = Horizontal Movement of Work (M)
	(H2 is higher than H1)	t = Element Time (Min.)

If T_{rest} is divided by T_{job}, the result is proportion rest time on job, and is independent of length of shift, given that E_{pwc} has already been assigned. Figure 1 plots the proportion of rest required as a function of E_{work} for two different PWC's. Note that the aerobic capacities of the average male and female, respectively, are 16 and 12 Kcal/min, so the chosen E_{pwc} will likely fall between the two ranges shown, for 4 to 8 hour shifts. It remains to be determined whether the E_{pwc} should remain the same in microgravity as in 1-G.

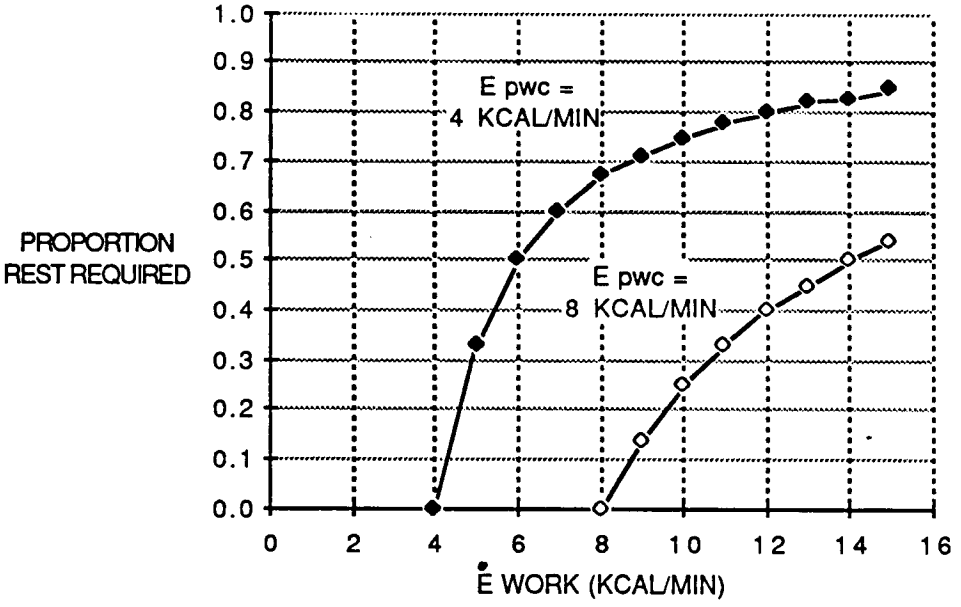


FIGURE 1. PROPORTION OF JOB SHIFT TIME SPENT RESTING AS A FUNCTION OF WORKING ENERGY EXPENDITURE. Plotted for two physical work capacities, based on Equation 2.

3. MODEL DEVELOPMENT: ENERGY EXPENDITURE PREDICTION IN REDUCED GRAVITY

When designing jobs for low gravity environments, it will be cumbersome to simulate every possible task on a low-gravity simulator, and physiologically measure oxygen consumption. Rather, empirical prediction equations are suggested as a way to ease this burden. While Garg's prediction model fits Earth-based industrial task data quite well, it is not known whether it can be used to predict lower gravity energy expenditures. A completely different set of relationships may indeed be in effect for walking, lifting, etc. On the other hand, weight reduction is likely the primary effect of lower gravity, as discussed by Wortz (1969), while mass acceleration characteristics will remain unchanged . By maintaining a constant basal metabolic rate, combined with decreased body weight and load weight in specific task energy contributions, estimates of workload may be made and compared with actual oxygen consumption data from low gravity simulations.[2] As a further simplifying assumption, the effect of a pressure suit and backpack may be estimated by adding their extra weight to an individual's body

[2]The assumption of a constant basal metabolic rate in lower gravity has been upheld by both Skylab and Apollo data (see Johnston and Dietlein, 1977; Johnston, et al.,1975).

weight in the task energy expenditure computations. The effects of pressure are not predicted by such model computations.

In the following sections, energy expenditures from specific tasks are reported, then predicted using Garg's (1976) model. For this analysis, work is divided into lower body work (walking), upper body work (stationary pushing and pulling), and whole body work (cart pushing and pulling). Table 3 provides an overview of the 8 studies in which sufficient task descriptions existed to allow use of the prediction model described above. Note that the relatively few subjects used in each study is not unusual for physiological studies. The average weight of these subjects varied from about 67 to 85 Kg, a fairly homogeneous group. All were healthy males, aged 20-40. Either inclined plane or vertical suspension systems were used to simulate low gravity (except, of course, in the Lunar EVA report), and some studies used pressure suits, pressurized to operational requirements. The lower-body work studies used walking speeds and walking slopes as independent variables. The upper-body work study measured torquing ability at a constant repetition rate. The whole-body work study measured ability to pull a small cart at varying velocities, weights, and slopes. Clearly, many other examples of whole-body tasks exist, but this was the only controlled study reported in the literature.

3.1. Lower Body Work

Studies of energy expenditure during 1/6 G walking were made during the latter 1960's, in preparation for the Apollo moon landings. Primary emphasis in these studies was placed upon differences in pressure suits, allowed walking speed, and task demand on the Lunar surface. One's E dramatically increases with speed of walking in either 1 G or 1/6 G environments. Wortz (1969) combined data from several sources to produce the observed relationship shown in Figure 2 (see Wortz et al. ,1971; Wortz and Prescott ,1966; Robertson and Wortz ,1968; and Passmore and Durnin ,1955). Observed E values, shown as a solid line, ranged between a resting 1.5 Kcal/min and a very difficult 11 Kcal/min for the treadmill study data. Observed E in 1/6 G was 1/4 to 1/2 that of E in 1 G, at all velocities. Using the mean body weight of 70 Kg from these studies, predicted E's were computed and plotted as a dashed line on the same figure. The computed values overpredicted observed 1 G data by 1 to 2 Kcal/min, but were much closer at the lower gravity level than at 1-G. This figure also displays 1/3 G predicted E's, as an estimate of walking energy expenditure on Mars.

The increased difficulty of walking with increasing velocity at 1/6 G was subjectively confirmed by Armstrong (1970). His impressions were that walking speed is limited to about 2 feet/sec (2.2 Km/hr) because "...The force at the foot necessary to maintain the speed will provide sufficient upward force to lift the person off the ground before the other foot comes down." This technique, called loping, is much like running in slow motion, in that both feet leave the ground at the same time. Steady state loping velocities were commonly 3-5 feet/sec (3.3-5.5 Km/sec) on the Lunar surface. During the 1/6 G simulations, subjects were forced to walk, rather than lope,

and the data may reflect E values that are excessive. Decreasing observed E's would, of course, increase the prediction error of Garg's model.

TABLE 3. LOW-GRAVITY ENERGY EXPENDITURE STUDIES

REFERENCE	NO. SUBJS	AVE. WT. (KG, 1G)	SIM*	CLOTHING†	SIMULATED GRAVITY (G)	EXPERIMENTAL CONDITIONS
			LOWER-BODY WORK			
Robertson and Wortz (1968)	6	71.0	VS, IP	PS	1, 1/6	1, 2, 4 mph
Wortz and Prescott (1966)	9	80.7	VS	SS	1/4, 1/6, 1/8	2, 4 mph
Sanborn and Wortz (1967)	10	72.5	IP	SS	1/6	2, 4 mph
Wortz (1969)	1	75.0	VS	SS	1, 1/2, 1/4, 1/6	4 mph
Wortz, et al., (1969)	6	67.7	VS, IP	SS, PS	1/6	2, 4, 6, 8 kph 0, 7.5, 15, 30°
Waligora and Horrigan (1975)	2	78.0	LS	PS	1/6	0.7-5.7 kph -13.8 - 11.3°
			UPPER-BODY WORK			
Prescott and Wortz (1966)	7	84.9	VS	SS	1, 1/2, 1/6, 0	Push: 2/second Push/pull: 2/s
			WHOLE-BODY WORK			
Camacho, et al. (1971)	2	74.8	VS	PS	1/6	75,148 Kg carts, 1,2,3,4,5 kph -15, 0, 15° slopes

*Simulator Codes: IP=Inclined Plane, VS=Vertical Suspension, LS=Lunar Surface EVA
†Clothing Worn During Study: SS=Shirt Sleeve, PS=Pressure Suit

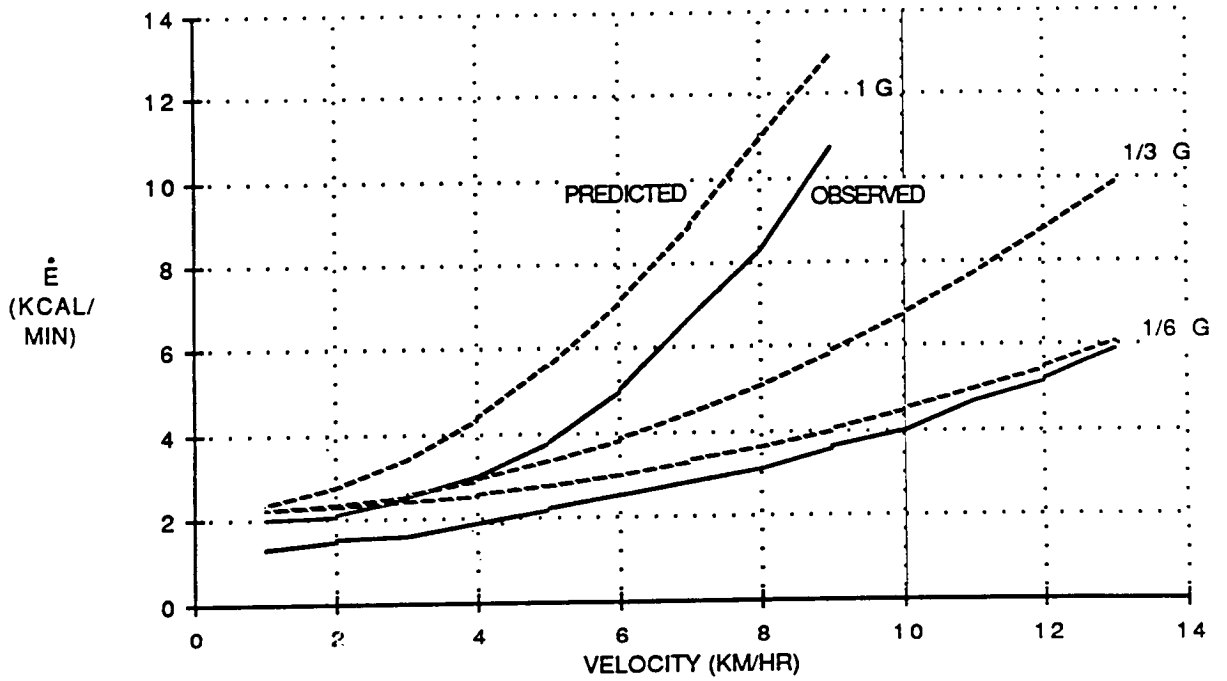


FIGURE 2. ENERGY EXPENDITURE AS A FUNCTION OF WALKING VELOCITY. Observed results (solid line) adapted from Wortz (1969), who summarized four treadmill studies. Predicted expenditures (dashed line) are from Garg's (1976) model.

Traction between a sole and the ground is linearly dependent on weight, so traction on the Moon is 1/6 that of an equivalent task on Earth. To successfully predict low gravity work, it is thus necessary to assess how much traction plays a part in increasing otherwise low walking energy expenditures in low gravity. This was done by Wortz (1969) by measuring the energy expended while maintaining a 1.79 Km/hr (4 Mi/hr) walk on a treadmill under gravity conditions between 1 and 1/6 G. Weight, and subsequent energy expenditure were reduced as simulated gravity decreased, shown in the lower curve of Figure 3.[3] The upper curve of this figure illustrates E values as weight was added to compensate for weight decreases due to simulated gravity. Thus, the traction and weight were the same at all tested gravity levels for the upper curve. The observed E only increased by about 1 Kcal/min between 1 and 1/6 G, even though the subjects were carrying an extra load of about 60 Kg. The observed rise in E was likely

[3]Percentage changes in E as shown by Wortz were converted to Kcal/min by using 7.0 Kcal/min as a baseline at a 1.79 Km/hr treadmill velocity, from Wortz and Prescott's (1966) analysis. The average 1-G body weights were within 5 Kg between these studies, so the estimated E is likely close to the actual E.

due to the effects of having to accelerate this extra mass with each step, as walking velocity is maintained. Indeed, weight seems to account for most of the decrease in E with decreasing gravity. Using the Walk element, with a constant E_{basal}, the solid data points in Figure 3 represent the predicted E values. At 1 G, predicted E was about 1.5 Kcal/min greater than actual, which could have been due to subjects of higher than average fitness level. Of course, the prediction model cannot adequately predict the increasing E at low G values due to increased load acceleration. The model very closely predicted the E declines at the three simulated gravity levels when no extra load was applied to the body.

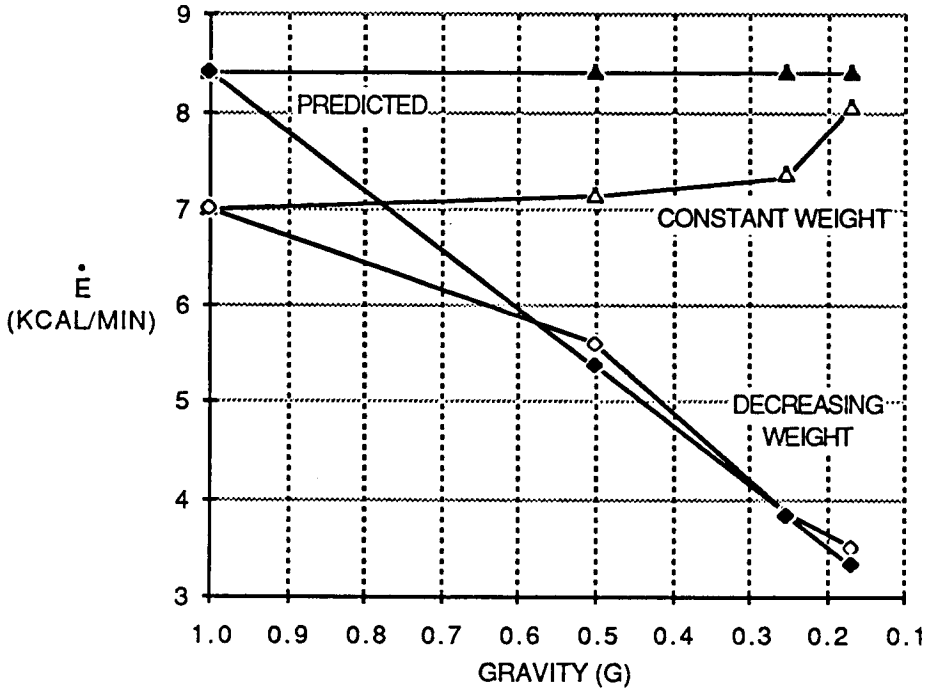


FIGURE 3. ENERGY EXPENDITURE AS A FUNCTION OF GRAVITY LEVEL. Adapted from Wortz (1969), by substituting 1-G expenditures from Wortz and Prescott (1966), as described in text. This data shown as light symbols, while predicted data are indicated by dark symbols.

Energy expenditures were obtained from the rate of oxygen consumption and heart rate during Lunar EVAs by Edwin Mitchell and Alan Shepard on Apollo 14. Relatively controlled observations of walking velocity, grade, and weight carried were used to predict E values. Two patterns emerged from this data, as shown in Figure 4. When climbing slopes steeper than 5%, E's climbed dramatically to the point where the workload was unduly heavy, when compared with slopes of less than 5%. When slopes were more reasonable, a relationship similar to that found in earlier simulations was obtained.[4] This EVA data was predicted by incorporating the appropriate slope,

velocity, etc. information into Garg's model. In general, the model closely predicted the simulator data, which underpredicted the observed data more and more as velocity increased. This error will be discussed in the following section.

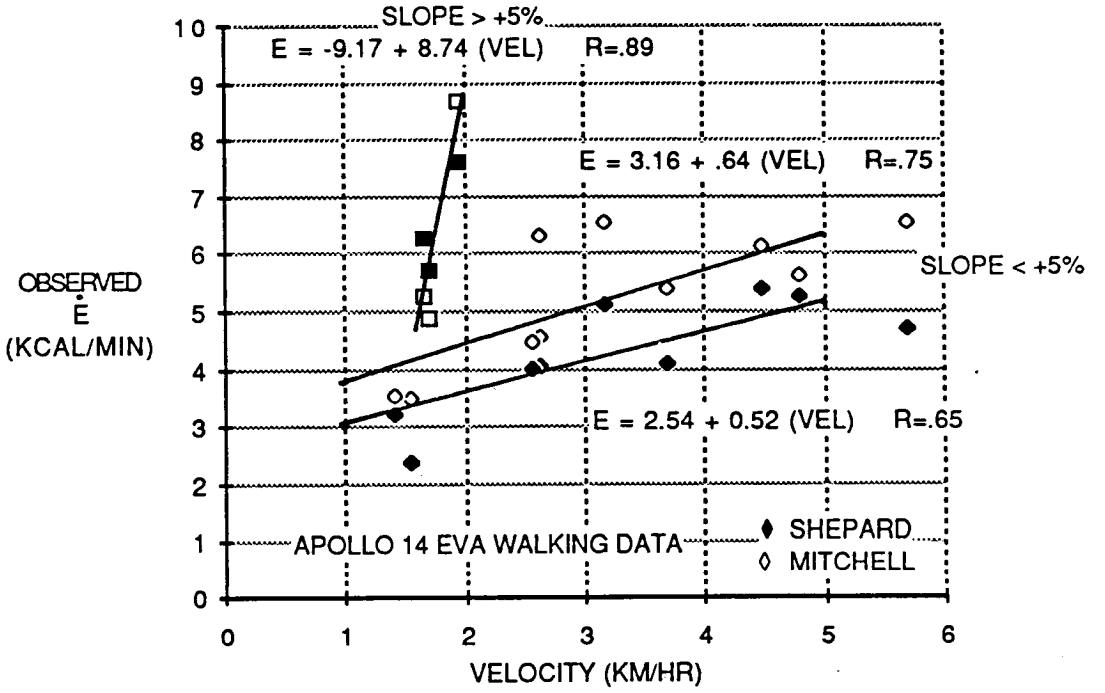


FIGURE 4. OBSERVED ENERGY EXPENDITURES DURING APOLLO 14 EVAS. Source: Waligora and Horrigan (1975). Shepard (dark data) and Mitchell (light) showed similar relationships between E and walking velocity.

3.2. Lower Body Prediction Error Analysis

An analysis of prediction error may be carried out by plotting observed versus predicted energy expenditures. Figure 5 shows this relationship for simulated gravity data. The data from Wortz, et al. (1969) agreed most closely with predicted values, with the model underpredicting observed data by an average of 15%. Note that this data was from subjects in shirt sleeves, and on horizontal ground. On the other hand, it was obtained by combining data from four studies. An extremely good prediction, accounting

[4]Though Lunar E's were 2.3 times higher than equivalent condition simulation E's, the simulated data was in fact highly predictive of the Lunar data. A linear regression between Lunar data from the average of Mitchell and Shepard's regression functions in Figure 4, and the simulated 1/6-G walking data from Figure 2 was: Lunar E = 2.13 + 1.55 (Simulated E). This relationship was quite linear (R=.98), and could be considered a 'real-world' correction factor, accounting for differences in the soil conditions, and the use of pressured suits.

for more than 95% of data variability, was obtained in these limited conditions. When predicting more isolated data, based on fewer subjects, the model may greatly underpredict energy expenditures, as shown by other data (squares)by Wortz and Prescott (1966). Wortz's (1969) 1/6-G observation was predicted very closely by the model.

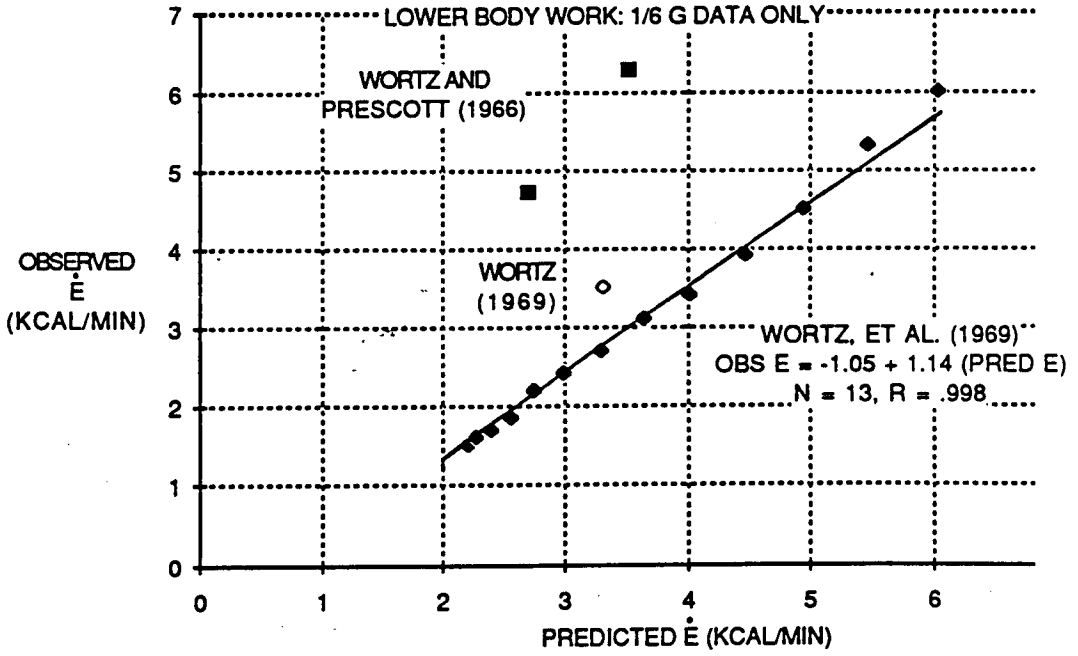


FIGURE 5. ENERGY EXPENDITURE PREDICTION ERROR FOR SIMULATED 1/6-G, TREADMILL DATA. Obtained by predicting E values for reported E's in given conditions.

Data from Lunar EVAs was also subjected to prediction error analysis, as shown in Figure 6. The Lunar data was obtained over many walking velocities, net slope changes, and surfaces which likely added to prediction error. A linear relationship between observed data and prediction model still existed (R=.65), but the mean observed energy expenditure value was underpredicted by about 30%.[5] The model so greatly underpredicted these values when slopes were greater than +5% that this data was not included in Figure 6. It also must be kept in mind that only two subjects were represented in the EVA data, clearly increasing prediction variability.

---

[5]This underprediction, due to unaccounted factors, is in line with the previously reported simulator-real world difference of 30-50%. The close agreement between the prediction model and simulator data underscored the fact that a 'correction factor' of 2-4 Kcal/min should be added when attempting to predict actual pressure-suited, EVA energy expenditure on the Lunar surface.

3.3. Upper and Whole Body Work

Most work requires the use of the upper torso, in addition to mobility gained by use of the legs. Prediction of energy expenditures in these tasks requires other than just walking or carrying elements. Use of the hold and push/pull elements are examples of additional prediction equations that must be used. It is important to note that Garg's (1976) model was not developed for continuous pushing or pulling of heavy loads over extended distances. Thus, much of the analysis below represents extreme extrapolations.

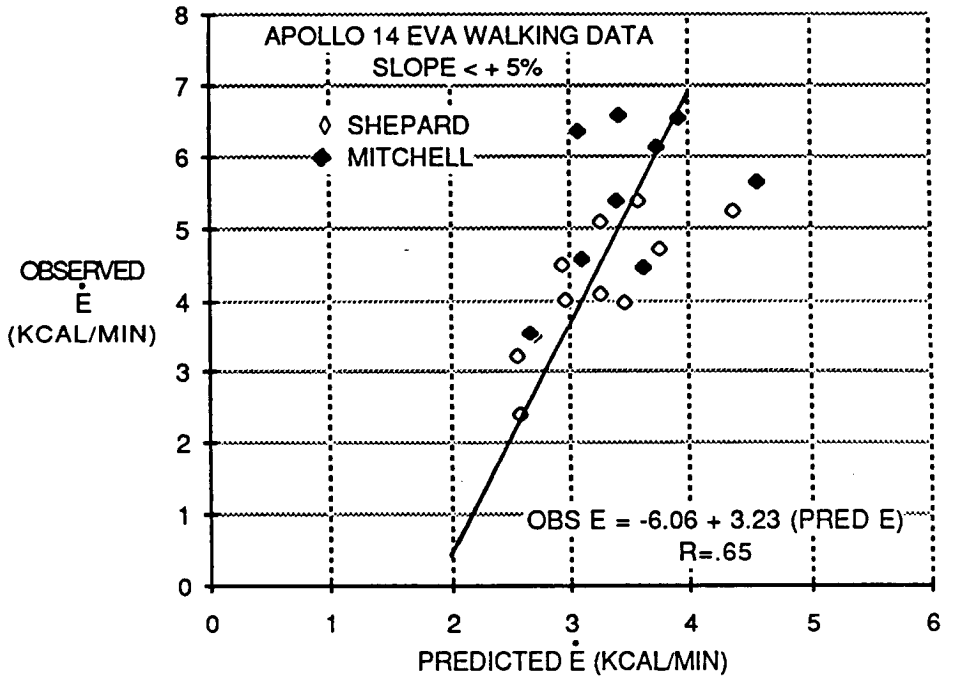


FIGURE 6. ENERGY EXPENDITURE PREDICTION ERROR FOR LUNAR EVA DATA. Lunar EVA data taken from Waligora and Horrigan (1975), using data only from net slopes of less than +5%. Prediction error from slopes greater than +5% was extremely large.

Prescott and Wortz (1966) predicted that loss of traction with lowered gravity levels should greatly decrease one's ability to torque bolts or perform other upper body work. They simulated a 9.2 M-Kg (800 in-lb.) torquing task, using simulated G levels of 1, 1/2, 1/6, and 0. Measured and predicted energy expenditures are shown in Figure 7. Two types of torquing tasks were used: a Push-only task where subjects torqued forward then returned to rest position, at a rate of 2/second. A Push/Pull task required torquing in both directions at the same rate. The two tasks produced essentially equivalent E values of 3-4 Kcal/min. As gravity was reduced, E's did not appreciably increase until weightlessness was achieved. Garg's model predicted E's nearly twice as

great as the observed values, using the Push at element. Body weight did not enter into this equation, so constant E was predicted with lower gravity. Garg's model thus needs to take tractional changes into account with lowered gravity. The model also utilizes the same work element for pushing and pulling, so only one set of predicted data are plotted.

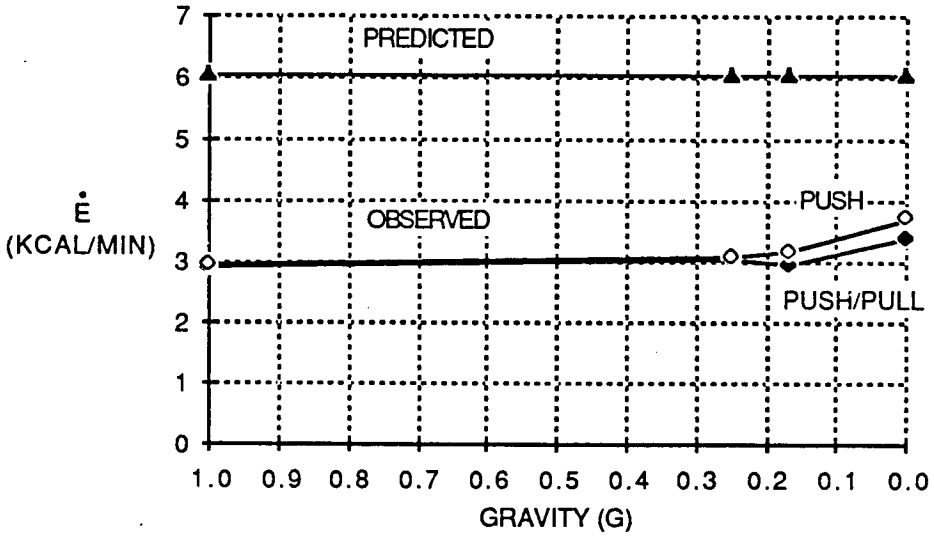


FIGURE 7. ENERGY EXPENDITURE DURING UPPER BODY TORQUING. Source: Prescott and Wortz (1966). Predicted E values (triangles) are from Garg's (1976) model. Both Push and Push/Pull tasks were performed at 2/second, and in shirt sleeves.

An analysis of whole body energy expenditure was provided by the cart pulling studies of Camacho, et al. (1971). In preparation for using a small cart on the Lunar surface, two cart weights (75, 148 Kg), five pulling velocities (1-5 Km/hr), and 3 walking grades (-15, 0, 15°) served as independent variables. Figure 8 provides observed and predicted E's for the conditions tested. For very steep, 15° slopes, observed E's were greater than 8 Kcal/min, and increased by 4 Kcal/min between 1 and 2 Km/hr. The prediction model does not specifically cover cart pulling, so the hand push/pull element was utilized, by inserting the equivalent hand force (as reported by Camacho, et al., 1971) exerted over a distance of 1 meter. This energy expenditure was then converted to a rate by correcting for velocity of movement. The model underpredicted the steep, 15° slope E's, but was fairly accurate on 0° and -15° slopes. On the steep downhill, energy expenditures were quite a bit smaller than on the other slopes, and this was very closely predicted. Cart pulling on level ground increased with velocity, and the model overpredicted the light cart E's by only 1 Kcal/min. The heavy

cart E's were also overpredicted, but by 1-4 Kcal/min, with error increasing as velocity increased. At 2-4 Km/hr, the heavy cart only produced 1-2 Kcal/min greater expenditure than the light cart.

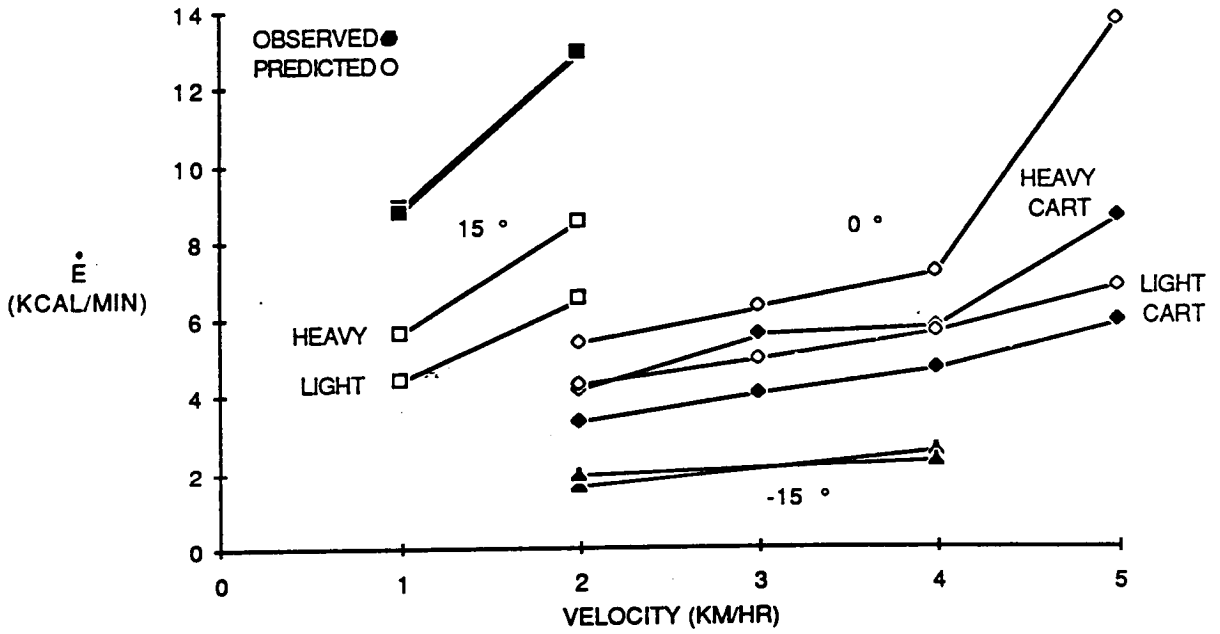


FIGURE 8. ENERGY EXPENDITURE DURING CART PULLING. Source: Camacho, et al. (1971). Predicted E values, from Garg's (1976) model are indicated by light data points.

3.4. Whole Body Prediction Error Analysis

The torquing data by Prescott and Wortz (1966) was overpredicted by 3 Kcal/min at nearly all tested gravity levels, and did not account for increasing E due to decreasing foot traction. The reason for the large overprediction is likely due to the fact that the tested subjects were of much higher fitness level than the average, healthy male adult working population. The torquing task, repeated twice per second, should have been more fatiguing. The investigators did not find any change in basal energy expenditure as gravity level was reduced, so the greatest difference in prediction was due to the task energy expenditure.

The cart-pulling studies provided more hope for prediction of whole-body energy expenditures. Observed versus predicted data from all conditions is shown in Figure 9. The excessively high E values for 15° slopes had a qualitatively different observed versus predicted relationship than the horizontal or downward slopes. A linear regression fit to the latter data clearly showed that cart pulling data can be well

predicted (R=.97) by Garg's (1976) model. Conditions were very realistic in this study, in that pressure suits were worn, and pulling took place on a Lunar soil simulant. Thus, it appears that the model can account for pressure suits and simulator conditions, but has a difficult time with steep upward slopes. A rough indication of 15° slope E values could be obtained by adding 4-6 Kcal/min onto predicted values.

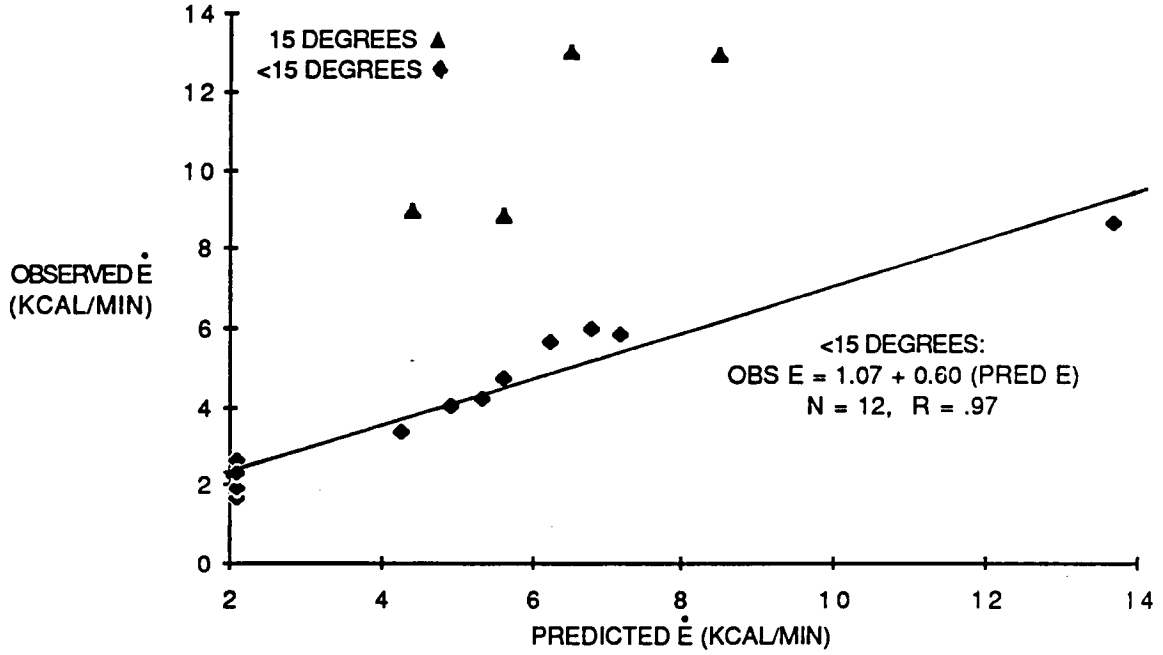


FIGURE 9. ENERGY EXPENDITURE PREDICTION ERROR FOR LUNAR CART PULLING. Relationship for 15° slope (triangles) was qualitatively different from -15° and 0° slopes (diamonds).

4. MODEL APPLICATION: WORKLOAD ANALYSIS OF CORE SAMPLING IN LOW GRAVITY

Blacic, Rowley, and Cort (1985) provided a conceptualization of rock drilling on Mars or the Moon. A portable jack-hammer drill will be connected, via high-pressure tubing, to a compressed CO_2 tank, on skids. The drill will have a depth capability of 2-3 meters, drilling a 50 mm diameter hole, into hard rock, in 30 minutes. The CO_2 will provide oscillatory motion to the drill bit, and will clear out the hole as drilling continues. The 1-G weight of the drill and skid, scaled from present equipment, were estimated to be 90 Kg.

To estimate the physical workload of a drilling outing, a number of assumptions must be made. A 70 Kg male, wearing a 20 Kg suit and 30 Kg life-support pack, will perform the drilling. An average surface grade of 5% will be negotiated; this parameter will

compensate for small surface irregularities. The drill is driven to a core site via a rover, where the unit is riding .80 M off the ground. Upon arriving at the site, the worker grabs the drill and skid and carries them to the site of the first hole, where they are lowered. All walking and carrying is carried out at .56 M/sec (2Km/hr), and successive core samples are collected 200 meters apart. Drilling is carried out by holding the operating drill for 30 minutes. The CO_2 unit is lifted and carried to the next drilling site, and the next hole is drilled. Five cores are drilled in this manner, by travelling around the rover. After all samples are collected, the worker will return to the rover for driving back to base. The physical work aspect of the shift, from leaving the rover to returning to the rover, requires 186 minutes. Analysis of this activity provided the data shown in Table 4. From Table 1, this task would be considered moderate work in 1-G, Light work in 1/3-G, and Very light to Light work in 1/6-G. The greatest energy expenditure in this job, besides basal, was due to carrying the drill 200 meters in between core samples. Due to the low frequency of task repetition, lifting and lowering accounted for a small percentage of overall energy expenditures. The contribution of lifted load to overall E is relatively small in lower gravity environments. Subsequent analysis has shown that, for a 50 Kg load increase (to 140 Kg), the associated increases in E for this job are 37%, 20%, and 10%, for 1-G, 1/3-G, and 1/6-G, respectively. A physical work capacity of 4 Kcal/min could be presumed if this task were carried out for 8 hours. If this were the case, 4 hours of rest would be required on Earth, and no additional rest on Mars or on the Moon. It is likely that the PWC is greater than this, however, due to the 3 hour work shift. Thus, no additional rest need be added to this core sampling job.

TABLE 4. PREDICTED ENERGY EXPENDITURES FROM CORE SAMPLING JOB

	PREDICTED ENERGY EXPENDITURES PER 186 MINUTE SHIFT					
LOCATION	BASAL	CARRY	LOWER	HOLD	LIFT	TOTAL
Earth (1-G)	2.00	2.96	0.02	1.13	0.05	6.16
Mars (1/3-G)	2.00	1.07	0.01	0.38	0.02	3.48
Moon (1/6-G)	2.00	0.60	0.01	0.19	0.01	2.80

5. CONCLUSIONS AND RESEARCH NEEDS

This paper has outlined a methodology for predicting energy expenditures for physical work in low gravity environments. It must be emphasized again that the form of these models may change quite a bit for lower gravity levels, and only empirical study can elicit these differences. Nevertheless, some E values from studies were surprisingly well-predicted, while others were not. The single element of walking was well predicted both in simulators and on the Lunar surface, so long as slopes were relatively small. The lunar data also required an additional 2-4 Kcal, due to pressure

suits and other real factors. Upper-body torquing was overpredicted by 3 Kcal/min at nearly all gravity levels, while cart pulling was accurately predicted at horizontal and downhill slopes. The latter was promising, as Garg did not include cart pulling in his prediction equations.

To obtain accurate predictive low gravity energy expenditure models, extensive simulation under highly controlled conditions is required. This paper has demonstrated the feasibility and use of such equations, but much is left to be done. More predictive equations are required for tasks such as pushing and pulling objects up and down grades, loping, digging, etc. The definition of shift rest time is completely dependent on Earth-based concepts of physical work capacity. It is highly likely that one's PWC increases at lower gravity levels, but this must be shown.

6. REFERENCES

American Industrial Hygiene Association (AIHA) (1971), Ergonomics Guide to Assessment of Metabolic and Cardiac Costs of Physical Work, Akron, OH: AIHA.

Armstrong, N. (1970), "Lunar Surface Exploration," in Kondrat'ev, K., Rycroft, M.J., and Sagan, C. (Eds.), Space Research XI: COSPAR, Plenary Meeting and Symposium on Remote Sounding of the Atmosphere Proceedings, Vol 1, International Union of Geodesy and Geophysics and the World Meteorological Organization.

Ästrand, P.O., and Rodahl (1977), Textbook of Work Physiology, 2nd Ed., McGraw-Hill.

Bink, B. (1962), "The Physical Working Capacity in Relation to Working Time and Age," Ergonomics, 5: 25-28.

Blacic, J.D., Rowley, J.C., and Cort, G.E. (1985), "Surface Drilling Technologies for Mars," in Manned Mars Missions Working Group Papers, Workshop at Marshall Spaceflight Center, Huntsville, AL, NASA M0002, June 1986, pp. 458-469.

Camacho, A., Robertson, W., and Walther, A. (1971), "Study of Man Pulling a Cart on the Moon," NASA CR-1697, Contractor Report, AiResearch Manufacturing Co.

Garg, A. (1976), A Metabolic Rate Prediction Model for Maual Materials Handling Jobs, Unpublished Doctoral Dissertation, The University of Michigan.

Garg, A., Chaffin, D.B., and Herrin, G.D. (1978), "Prediction of Metabolic Rates for Manual Materials Handling Jobs," American Industrial Hygiene Association Journal, 39:661-674.

Grandjean, E. (1981), Fitting the Task to the Man, New York: Int. Publications Service.

Johnston, R.S., Dietlein, L.F., and Berry, C.A. (1975), Biomedical Results of Apollo, National Aeronautics and Space Administration, NASA-SP-368.

Johnston, R.S., and Dietlein, L.F. (1977), Biomedical Results from Skylab, National Aeronautics and Space Administration, NASA SP-377.

National Commission on Space (1986), Pioneering the Space Frontier, Bantam Books: Toronto.

Passmore, R., and Durnin, J.V.G.A. (1955), "Human Energy Expenditure," Physiological Reviews, 35: 801-875.

Prescott, E.J., and Wortz, E.C. (1966), "Metabolic Costs of Upper Torso Exercises vs Torque Maneuvers Under Reduced-gravity Conditions," Aerospace Medicine, 10: 1046-1049.

Robertson, W.G., and Wortz, E.C. (1968), "Effect of Lunar Gravity on Metabolic Rates," Aerospace Medicine, 8:799-805.

Sanborn, W.G., and Wortz, E.C. (1967), "Metabolic Rates During Lunar Gravity Simulation," Aerospace Medicine, 4: 380-382.

Sanders, M.S., and McCormick, E.J. (1987), Human Factors in Engineering and Design, 6th Ed., McGraw-Hill Co.: New York.

Waligora, J.M., and Horrigan, D.J. (1975), "Metabolism and Heat Dissipation During Apollo EVA Periods," in Johnston, R.S., Dietlein, L.F., and Berry, C.A., Biomedical Results of Apollo, National Aeronautics and Space Administration, NASA-SP-368.

Webb, P. (1973), "Work, Heat, and Oxygen Cost," in Parker, J.F., and West, V.R. (Eds.), Bioastronautics Data Book, 2nd Ed., National Aeronautics and Space Administration.

Wortz, E.C. (1969), "Work in Reduced-Gravity Environments," Human Factors, 11(5):433-440.

Wortz, E.C., Robertson, W.G., Browne, L.E., and Sanborn, W.G. (1969), "Man's Capability for Self-Locomotion on the Moon, Volume II-Summary Report," AiResearch Report No. 68-4262, Rev. 1, and NASA-CR-1403.

Wortz, E.C., and Prescott, E.J. (1966), "Effects of Subgravity Traction Simulation on the Energy Costs of Walking," Aerospace Medicine, 12: 1217-1222.

A SOLID PHASE ENZYME-LINKED IMMUNOSORBENT ASSAY FOR THE ANTIGENIC DETECTION OF LEGIONELLA PNEUMOPHILA (SEROGROUP 1): A COMPLIMENT FOR THE SPACE STATION DIAGNOSTIC CAPABILITY

NASA/ASEE Summer Faculty Research Fellowship Program--1987

Johnson Space Center

Prepared by:	Kelly E. Hejtmancik, Ph.D.
Academic Rank:	Instructor
University & Department:	Galveston College
	Division of Mathematics and Science
NASA/JSC	
Directorate:	Space and Life Sciences
Division:	Medical Sciences
Branch:	Biomedical Laboratories
JSC Colleague:	Duane L. Pierson, Ph.D.
Date:	August 7, 1987
Contract#	NGT 44-001-800 (Texas A&M University)

ABSTRACT

A SOLID PHASE ENZYME-LINKED IMMUNOSORBENT ASSAY FOR THE ANTIGENIC DETECTION OF *LEGIONELLA PNEUMOPHILA* (SEROGROUP 1): A COMPLIMENT FOR THE SPACE STATION DIAGNOSTIC CAPABILITY

Kelly E. Hejtmancik, Ph.D.
Instructor
Division of Mathematics and Science
Galveston College
Galveston, Texas 77550

It is necessary that an adequate microbiology capability be provided as part of the Health Maintenance Facility (HMF) to support expected microbial disease events and environmental monitoring during long periods of space flight. The applications of morphological and biochemical studies to confirm the presence of certain bacterial and fungal disease agents are currently available and under consideration. This confirmation would be facilitated through employment of serological methods to aid in the identification of bacterial, fungal, and viral agents.well. A number of serological approaches are currently being considered, including the use of Enzyme-Linked Immunosorbent Assay (ELISA) technology, which could be utilized during microgravity conditions.

A solid phase, membrane supported ELISA for the detection of *Legionella pneumophila* (Serogroup 1), an expected disease agent, was developed to show a potential model system that would meet the HMF requirements and specifications for the future space station. These studies demonstrate the capability of membrane supported ELISA systems for identification of expected microbial disease agents as part of the HMF.

NASA Colleague: Duane L. Pierson, Ph.D. SD4 X3-7166

INTRODUCTION

The health and well being of individuals aboard a space station and possibly during future long space missions is of priority and must be assured. Certain expected clinical syndromes and diseases have been identified through an infectious disease conference conducted at the Johnson Space Center during October, 1985. Legionella pneumophila was identified as a possible problem due to its tendency to grow in filtration and water systems. Since 1976, 23 Legionella species comprising 37 serotypes have been identified. L. pneumophila, the most prevalent Legionella species in the United States, currently contains 12 serogroups, all which have been involved in pneumonia in humans (8). Previous spaceflight studies indicate a high probability of cross-contamination among crew members and microbial build up of space modules during long confinements, such as minimal (90 day) missions which are planned for the space station (7). Continual habitation, crowded conditions, possible immunosuppression, and other factors may create critical situations aboard the space craft. If a microbial disease is suspected, the major effort would be directed toward obtaining some indication of the specific kind of microorganism causing the problem. The exact nature of the etiological agent would determine the severity of the disease, treatment, prophylaxis, and subsequent health measures for the space station environment.

The diagnosis of a microbial disease currently rests upon one of a combination of clinical signs and symptoms, morphological and biochemical identification of isolates, and/or serological procedures. Special procedures such as cell culture may also be required. One problem with

limiting the scope of diagnosis to clinical signs and symptoms is that a particular microbe can sometimes produce infection having very different clinical characteristics and occurring in widely different areas of the body. For example, antibiotic-resistant Staphylococcus aureus may produce skin and subcutaneous tissue lesions as well as pneumonia, osteomyelitis, bacteremia, and acute membranous enterocolitis depending upon the means by which the organism gained entrance to the body, host resistance, antibiotic therapy, and other factors.

While the principle of one microorganism causing one clinical disease is often valid, there are many situations where this is not true. Indeed, pneumonias that are hardly separable clinically may be produced by several different kinds of bacteria and viruses. Correct diagnosis and treatment therefore heavily depend upon the abilities of the clinical laboratory.

Over the past few years, many new immunological methods have been developed which now provide the clinical laboratory with a large array of potentially valuable diagnostic tools. Antibodies and antigens labelled with radioisotopes or fluorescent dyes, or affixed to particulate materials, have been used extensively for immunodiagnosis over the past three decades. These methods do have disadvantages. Immunofluorescence, for example, usually depends upon a subjective assessment of end result, and the technique is frequently laborious. Radioimmunoassay requires expensive equipment and carries the risk of radioactive exposure and contamination. Moreover, the current methods for either technology are not applicable to microgravity. The concepts that antigen and antibody can be attached to a solid phase support yet retain immunological activity, and

that either can be coupled to an enzyme and the complex retain both immunological and enzymatic activity, led to the development of Enzyme-Linked Immunosorbent Assays (ELISAs). Antibodies and antigens have been shown to readily attach to plastic surfaces (such as polyvinyl or polystyrene) either by passive absorption or chemical conjugation, and still retain immunological activity. Antibodies and antigens have been linked to a variety of enzymes including glucose oxidase, peroxidase, and alkaline phosphatase. Table 1 indicates the positive factors for use of ELISA systems in diagnostic microbiology.

Table 1: Positive Factors for HMF Consideration of ELISA Systems.

- ✓ Low Cost
- ✓ Reagent stability
- ✓ Safety
- ✓ Sensitivity
- ✓ Specificity
- ✓ Reproducibility
- ✓ Ease of procedure
- ✓ Can be performed in poorly equipped laboratories
- ✓ No power requirements

It appears that the space station diagnostic capability will most likely require immunological testing applicable to the identification of microorganisms, particularly those that cannot be cultured or identified by standard laboratory techniques. In recent years, there has been increasing emphasis on accurate, reliable, and quick immunological procedures for the identification of many microorganisms and/or the immunological responses of

the host toward infection. Most current commercially available procedures have been developed for use in clinical laboratories and not designed for microgravity conditions which would be present in the space station environment (5). It appears and is reasonable that a number of some existing procedures, particularly solid phase immunoassays, could be modified in regard to uniformity and standardization for use aboard the space station. This project was designed to illustrate the concept of a solid phase, membrane supported ELISA for detection of L. pneumophila (Serogroup 1) to demonstrate the capability of ELISA systems for identification of expected microbial disease agents aboard the space station.

MATERIALS

Equipment: A 96 well Bio-Dot filtration apparatus (#170-6550) was obtained from Bio-Rad Laboratories, Richmond, Ca. 94801.

Buffers: A 20 mM Tris buffered saline (TBS), pH 7.5, was prepared by adding 4.84 g Tris (Bio-Rad) to 58.48 g NaCl, brought to a 2 liter volume with deionized water. The buffer was adjust to pH 7.5 with 1 M HCl.

Blocking Solution: Bovine serum albumin (BSA, #A-7030) was purchased from Sigma. A 3% BSA-TBS was prepared by adding 3 g of BSA to 100 ml of TBS.

Wash Solution: A wash solution containing 0.05% Tween-20 was prepared by adding 0.5 ml of Tween-20 (Bio-Rad) to 1 liter of TBS.

Antibodies: Antisera prepared in rabbits against L. pneumophila (strain Philadelphia 1 and OLDA) were obtained from Dr. Hazel Wilkinson, the Department of Health and Human Services, Center for Disease Control,

Atlanta, Ga. 30333. These antisera were pooled for coverage of all serogroup 1 subtypes (9). Horseradish perioxidase conjugated (HPR) goat anti-rabbit serum (#170-6500) was obtained from Bio-Rad Laboratories.

Nitrocellulose Membranes: Nitrocellulose membranes (#162-0117) with a pore size of 0.45 microns was purchased from Bio-Rad Laboratories.

Legionella Antigen: L. pneumophila antigen was prepared from an ATCC 3152 (serogroup 1) lyophilized culture vial (3). The lyophilized culture vial was broken and the lyophilized material was dissolved into 4 ml of Trypticase Soy Broth. Four 15 x 100 mm plates containing 25 ml of buffered charcoal yeast extract (BCYE) agar each of which was inoculated with 1 ml of the dissolved material. The plates were enclosed in a plastic container to prevent the agar from drying out and were incubated at 35 degrees centigrade for 48 hr. The cells were suspended from each agar surface in 3 ml of sterile 0.01 M phosphate buffered saline, pH 7.2, with a pasteur pipette into a 25 ml sterile conical tube. The conical tube containing the cell suspension was boiled for 1 hr to kill the cells. The killed cell suspension was centrifuged at 1600 x g for 30 min, the supernatant discarded, and the cells resuspended in 2 ml of 0.01 M phosphate buffer, pH 7.2, for each 0.1 ml of packed cells. One drop of a 1:1000 methiolate solution was added for each 2 ml of preparation. The stock solution was stored at 4 degrees centigrade for 10 days to allow for the release of soluble antigen from the cells. The suspension was centrifuged at 1600 x g and the supernatant used for assay development. The antigen preparation was subjected to the Lowry protein detection method indicating the antigen preparation concentration was approximately 0.025 mg/ml. Subsequent

calculations were determined from this estimate.

Stock Chromogenic Substrate Stain Solution: Two substrates were utilized for comparison. O-phenylenediamine (OPD, Abbott Laboratories) was prepared by dissolving 12.8 mg into 5 ml of citrate phosphate buffer, pH 7.2, containing 0.02% hydrogen peroxide. 4-chloro-1-napthol (4C1N, Bio-Rad) was prepared by dissolving 60 mg of 4C1N into 20 ml of ice cold methanol. Immediately prior to use, 0.06 ml of ice cold 30% hydrogen peroxide was added to 100 ml of room temperature TBS. The two solutions were mixed just prior to use.

Preparations Obtained for Specificity Studies: All antigen preparations tested were obtained from Difco Laboratories. These preparations included Salmonella O Poly A-I and Vi (#2364-47-2) which contains antigens from Group A, B, C1, C2, D2, E1, E2, E4, F, G, H, I, and Vi; Shigella Group A (#2100-50-1), A1 (#2101-50-3), B (#2102-50-2), C (#2103-50-0), C1 (#2104-50-0), C2 (#2105-50-9), and D (#2106-50-8); Pseudomonas aeruginosa antigen set (#3082-32-7); Streptococcus antigen set containing Groups A, B, C, D, E, F, and G (#2368-32-4).

Immunofluorescent Assay Kit: A MonAbrite Legionella Polyscreen kit was obtained from Serono Diagnostics, Inc., 100 Longwater Circle, Norwell, MA 02061 and used for comparison with results obtained by the ELISA. This kit recognizes 21 species and 33 serogroups of Legionella including Serogroup 1.

METHODS

The ELISA system utilized the Bio-Dot apparatus with mounted nitrocellulose paper. The procedure for assembly of the apparatus and

preparation of the nitrocellulose paper was provided by Bio-Rad Laboratories (1). Figure 1 illustrates the basic components of the system. Nitrocellulose paper was first soaked in TBS to ensure uniform

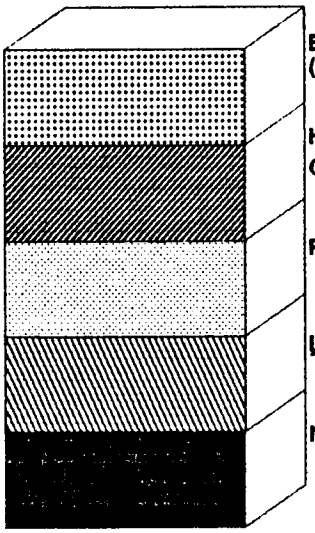


Figure 1: Major components of the ELISA for L. pneumophila.

protein bindings and low background absorption. The cleaned and dried Bio-Dot apparatus was assembled, and the nitrocellulose paper sheet wetted prior to being placed in the apparatus. The apparatus was appropriately tightened to insure that cross well contamination would not occur.

The flow valve was adjusted to allow the vacuum chamber to be exposed to the atmosphere and the appropriate wells to receive antigen preparations were inoculated with a 0.05 ml volume. The entire sample was allowed to filter through the membrane by gravity flow (approximately 30 min). Each well was filled with the same volume of sample solution to insure homogeneous filtration of all sample wells.

After the antigen samples completely drained from the apparatus, 0.2 ml of a 3% BSA/TBS blocking solution was applied to each well. Gravity filtration was allowed to occur until the blocking solution completely drained from each well (approximately 30 min).

The flow valve was adjusted to vacuum and 0.4 ml of wash solution (TBS with 0.05% tween) was added to each well. The wash solution was allowed to completely drain from all wells. This process was repeated. Following the wash step, the flow valve was opened to the atmosphere and 0.1 ml of the rabbit anti-Legionella antibody solution was added to each of the wells. The solution was allowed to completely drain from the wells, and another wash step performed.

With the vacuum off and the flow valve to the atmosphere, 0.1 ml of horseradish peroxidase labelled goat anti-rabbit solution was added to each well. The solution was allowed to completely drain from the wells.

Following the second antibody step, the vacuum was turned on and a wash step performed. Immediately, 0.2 ml of a color development solution, either OPD or 4C1N was applied to each well. A positive ELISA reaction was shown as color development depending upon the substrate utilized. The reactions were stopped by the addition of 0.3 ml of distilled deionized water to each well.

The MonAbrite Legionella Polyscreen (Serono Diagnostics, Inc) Immunofluroescent Assay Kit was accomplished by the procedure which accompanied the package. Briefly, test specimens and controls were applied as 0.05 ml aliquots to slides and heat fixed. Rabbit anti-Legionella was applied to each specimen with the exception of the control wells. After a

30 min incubation period, a wash procedure was performed. The slides were air dried and FITC-labelled goat anti-rabbit serum was applied to each specimen. Slides were washed, air dried, and examined for immunofluorescence utilizing a Nikon Episcopic-Fluorescence microscope with a B-1A filter cube.

RESULTS

Assay sensitivity: Dose response curves utilizing decreasing concentrations of L. pneumophila (Serogroup 1) antigen applied to the nitrocellulose membrane and titrated with increasing dilutions of L. pneumophila (Serogroup 1) antibody are shown in Figure 2. The reaction intensity was measured by eye from 0 to 4+, 4+ graded as the most intense of the color outcome. O-phenylenediamine was the substrate used in these experiments. As the antibody was diluted, the sensitivity of the ELISA was shown to decrease. Dilutions of antibody tested between 1:10 and 1:100 allowed the detection of L. pneumophila antigen to 0.0626 ng (62.6 pg); whereas, antibody dilutions of 1:500, 1:1000, 1:5000, and 1:10000 reduced the sensitivity of the ELISA to 0.3442, 0.626, 6.26, and 34.43 ng, respectively. Antibody dilutions including and exceeding 1:50000 did not allow for detection of L. pneumophila antigen. The optimal dilution of antibody for used in the ELISA was determined to be 1:100.

Assay specificity: Available antigen preparations from other bacterial groups were applied to the nitrocellulose membrane to test for cross reactivity with the anti-L. pneumophila serum. As indicated in Table 2, no cross-reactivity was detectable between the various groups or types of Salmonella, Shigella, Pseudomonas, or Streptococcus when these

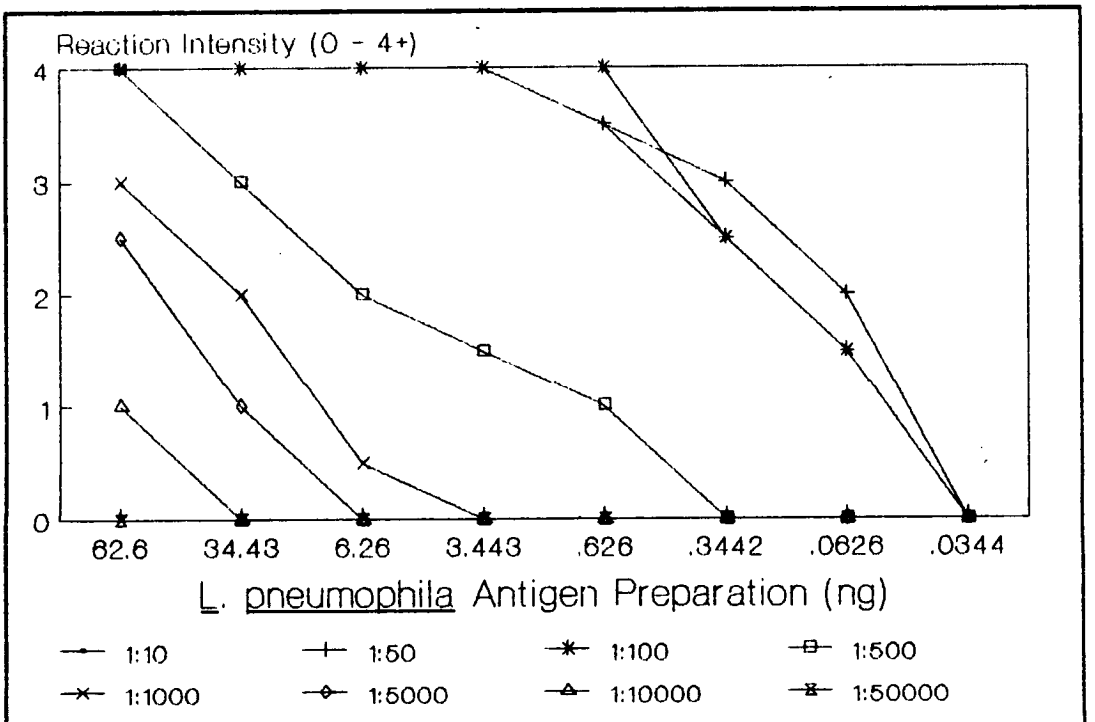


Figure 2: Titration of L. pneumophila (Serogroup 1) antigen and antibody preparations by ELISA methods. (Note: intervals are not equal units)

preparations were applied to the nitrocellulose membrane and tested by the ELISA. Each of the preparations were tested as 10x dilutions from the stock to a final dilution of 1:1000. Five urine, blood plasma, and two sputum samples were applied to the nitrocellulose membrane and were negative when tested by the ELISA. Dose response curves were not affected by either biological fluidwhen L. pneumophila was seeded into these specimens and applied to the nitrocellulose membrane.

Comparison of sensitivity between ELISA and IFA methods: The L. pneumophila antigen preparation was titrated by the MonAbrite Legionella Polyscreen kit (Serono Diagnostics, Inc.) and compared to that achieved by ELISA. The IFA results were based upon the brightness of field (0 to 4+)

Table 2: Antigens and Biological Fluids Testing Negative in the ELISA.

- Salmonella
  Groups A,B,C1,C2,D2,E1,E2,E4,F,G,H,I,Vi
- Shigella
  Groups A,A1,B,C,C1,C2,D
- Pseudomonas
  Types 1-17
- Stretococcus
  Groups A,B,C,D,E,F,G
- Urine
- Blood plasma
- Sputum

with relationship to the diluted antigen preparation. As indicated in Figure 3, ELISA results are similar to that indicated in Figure 2 for the standard dose response curve with minimum antigen detection between 0.0626 and 0.0344 ng. IFA results indicate a minimum detection limit of 0.626 ng of antigen, an approximate 10 fold reduction in sensitivity when compared with the results of the ELISA. A dilution curve of the polyvalent antigen control in the IFA kit was titrated by dilution using both ELISA and IFA techniques. In these determinations the ELISA was capable of detecting the control specimen to a 1:50 dilution; whereas, the IFA was readable to a 1:10 dilution. Since the Serono IFA kit is polyvalent and not specific toward Serogroup 1, strict interpretation of the comparsion of the

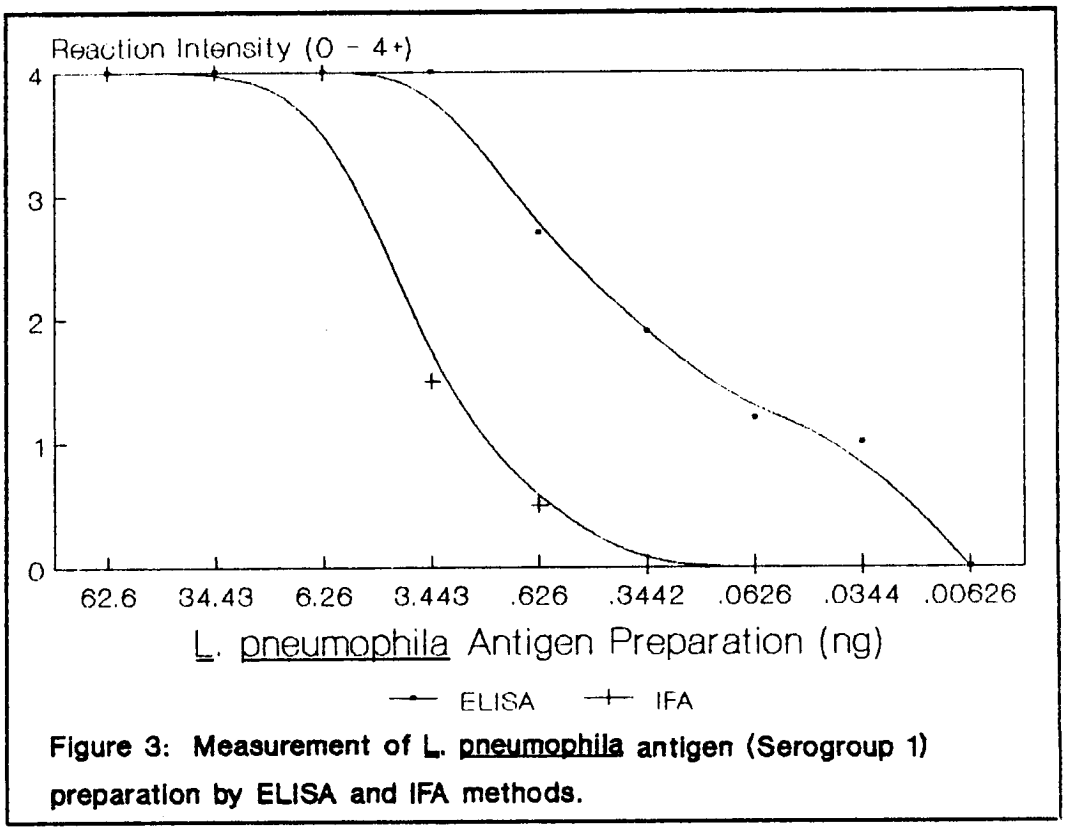


Figure 3: Measurement of L. pneumophila antigen (Serogroup 1) preparation by ELISA and IFA methods.

companion polyvalent antigen control between the two methods is not valid. However, in these determinations the ELISA was capable capable of detecting the control specimen to a 1:50 dilution; whereas, the IFA was readable to a 1:10 dilution.

Comparison of OPD and 4C1N as enzyme substrates: OPD (0-phenylenediamine) and 4C1N (4-chloro-1-napthol) were compared in the ELISA for their performance to detecting L. pneumophila antigen. As shown in Figure 4, the OPD was capable of detecting lesser amounts of antigen than 4C1N. Only 34.43 ng was detectable with 4C1N.

DISCUSSION

It appears logical that the methods selected for the diagnostic

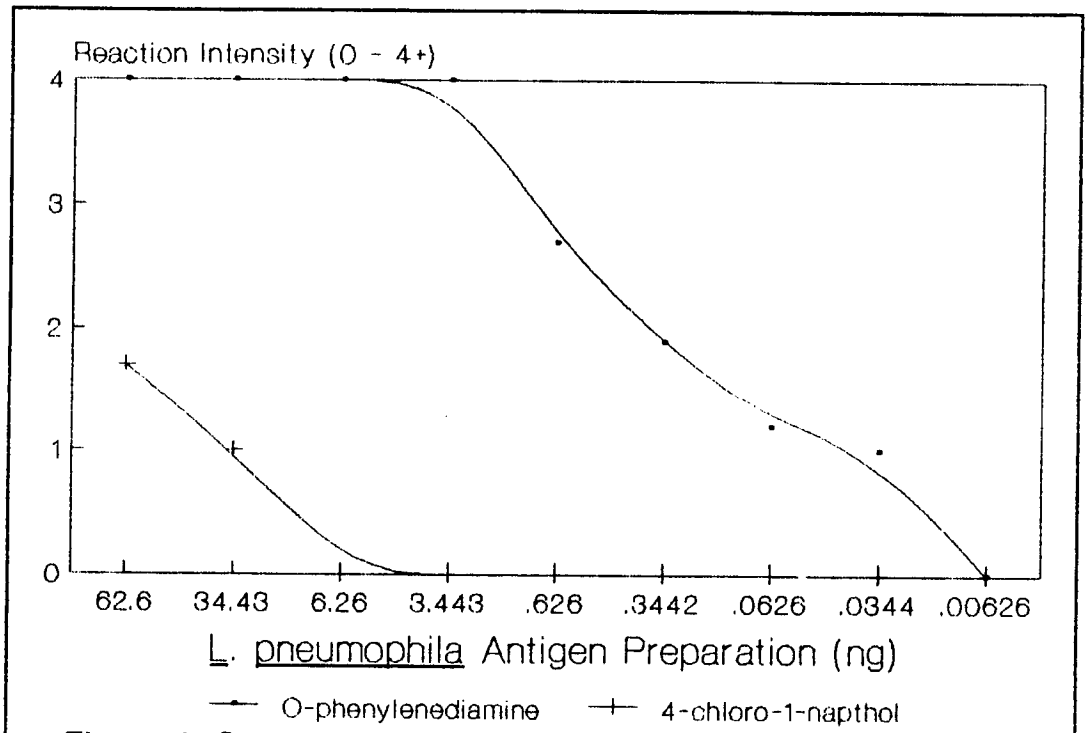


Figure 4: Comparison of OPD and 4C1N as Substrates in the ELISA.

systems Health Maintenance Facility (HMF) as well as microbiological environmental monitoring of space stations and vehicles will need to support identification of microorganisms that pose a health threat to the astronauts. A closed environmental system makes its necessary to monitor surfaces, water, air, waste systems, and possibly food for contamination with microorganisms. The growth of microorganisms in a closed recycled environment increases the potential for cross-contamination of crew members particularly when missions of 3 months and possible missions to Mars (1-3 years) are being considered. Depending upon the humidity and organic debris build-up in this environment, microorganisms may grow uninhibited. It is evident that the HMF and environmental monitoring systems must be in

place for microbial detection since quick return trips will be impossible. Thus, detection of a microbial problem could result in methods applied to eliminate that situation before more serious problems develop. Illnesses caused by microorganisms vary significantly in regards to symptoms, severity, prophylaxis, and treatment. As experienced during Apollo 13, a urinary tract infection with Pseudomonas aeruginosa could have resulted in a serious situation if the crew member had not been able to receive immediate health care (4). The potential for cross-contamination of crew members has been documented by chamber tests and experiences aboard the Skylab missions (2). These situations provide evidence that adequate microbiological diagnostic systems be in place aboard space vehicles or stations prior to their habitation for long time periods. There is currently considerable debate of on the effects of microgravity, radiation, temperature, and pressure on man, particularly immunosuppressive effects. If immunosuppression is indeed a reality during long space missions, the HMF must have the capability to not only provide identification of a microbial problem, but also medication against the invading microorganisms. During the past decade, numerous immunoassays have gained wide acceptance as the method of choice in the diagnosis of numerous disease states (6). The ideal considerations of a diagnostic test particularly for the NASA space program include speed, sensitivity, specificity, accuracy, safety, inexpensive reagents, potential for automation, long reagent shelf life, and broad applicability. Moreover, they must function in microgravity and not contain biologicals which would themselves threaten the health of the crew members. Although

many techniques have been recently developed for the immunological detection of antigens and/or antibodies, enzyme immunoassays such as the ELISAs are among the most popular both in research and clinical laboratories for the diagnosis of bacteria, protozoans, and viruses (5). In general, ELISAs are user-friendly, reliable, highly sensitive and specific, and require little time to run. Additional considerations include that no power source or instruments are required for the performance of the tests, little equipment is needed, reagents are stable, and results can be read by eye as intensive color changes.

The majority of commercially available ELISA systems are designed to detect antibody levels in blood plasma or other biological fluids such as urine and few have been developed for the detection of microbial antigens. At the present time, commercially available ELISA systems were not designed to be utilized in microgravity, and thus, little concern was given to designing these tests for those conditions. Only one ELISA, the Test Pack by Abbott Laboratories, released in June, 1986 for purchase, appears promising for space station use. This test detects the presence of Streptococcus pyogenes (Group A, beta-hemolytic streptococci) directly from throat swabs. The test performed well when tested aboard the KC 135 which simulates microgravity during parabolic curves.

Results from experiments conducted in this project and the existence of a commercially available kit paralleling these findings, provide a current technology to be considered for the HMF and environmental monitoring systems for space vehicles and stations. A major advantage to consider with the solid phase filter membrane systems is that the fluids

involved in the system can be retained resulting in little chance of spillage in the spacecraft environment. This project also presents an ELISA developed against Legionella pneumophila (Serogroup 1), the causative agent of Legionaires' disease. Because of its habitat, growth in filter and condensation units, this bacteria poses a possible health threat particularly in a closed recycled environment as that indicated aboard the space station and future space vehicles. The ELISA developed in this project could be utilized for monitoring and detecting its presence and levels in filters and condensates aboard these space crafts as well as to test a crew member for its presence in a respiratory disease syndrome.

REFERENCES

1. Bio-Dot Microfiltration Apparatus Instruction Manual. Bio-Rad Laboratories, Richmond, Ca., 1984.

2. Cintron, N.: JSC Microbiological and Toxicological Requirements for Space Station. Briefing Document, NASA/JSC Medical Sciences Division, Houston, Texas, August, 1985.

3. Farshly, C.E. and J.C. Feeley: ELISA for Legionaires' Disease - Antibody and Antigen Detection Systems: An Interim Report. In: "Legionaires'" the Disease, the Bacterium and Methodology. HEW Publication (CDC) #79-8375, 1979.

4. Furukawa, S.: Medical Support for Long Duration Space Missions. Applic. Space Technol., 3:203-209, 1983.

5. Hejtmancik, K.E.: Expansion of Space Station Diagnostic Capability to Include Serological Identification of Viral and Bacterial Infections. In: NASA/American Society for Engineering Education (ASEE) Summer Faculty Fellowship Program. B.C. McInnis and S.H. Goldstein (eds.), NASA Contractor Report #171984, June, 1987.

6. O'Beirne, A.J. and H.R. Cooper: Heterogeneous Enzyme Assay. J. Hist. Cyto., 27:1148-1157, 1979.

7. Pierson, D.L.: Microbiology Support Plan for Space Station. National Aeronautics and Space Administration Publication, JSC #32015, 1986.

8. Thacker, W.L., H.W. Wilkinson, R.F. Benson, and D.J. Brennex: Legionella pneumophila Serogroup 12 Isolated from Human and Environmental Sources. J. Clin. Microbiol., 25:569-572, 1987.

9. Wilkinson, H.W., B.J. Fikes, and D.D. Cruce: Indirect Immunofluoresence Test for Serodiagnosis of Legionaires' Disease: Evidence for Serogroup Diversity of Legionaires' Disease Bacterial Antigens and for Multiple Specificity of Human Antibodies. J. Clin. Micro. 9: 379-383, 1979.

POWER AND CHARGE DISSIPATION FROM AN ELECTRODYNAMIC TETHER

Final Report

NASA/ASEE Summer Faculty Fellowship Program—1987

Johnson Space Center

Prepared by:	Gerald E. Hite, Ph.D.
Academic Rank:	Associate Professor
University & Department:	Texas A&M University at Galveston Department of Marine Science Galveston, Texas 77553
NASA/JSC	
Directorate:	Space and Life Sciences
Division:	Solar System Exploration
Branch	Space Science
JSC Colleague:	James E. McCoy
Date:	August 1, 1987
Contract Number:	NGT 44-001-800

POWER AND CHARGE DISSIPATION FROM AN ELECTRODYNAMIC TETHER

The Plasma Motor-Generator project utilizes the influence of the geomagnetic field on a conductive tether attached to a LEO spacecraft to provide a reversible conversion of orbital energy into electrical energy. The behavior of the current into the ionospheric plasma under the influence of the geomagnetic field is of significant experimental and theoretical interest. Theoretical calculations are reviewed which start from Maxwell's equations and treat the ionospheric plasma as a linear dielectric medium. These calculations show a charge emitting tether moving in a magnetic field will generate electromagetic waves in the plasma which carry the charge in the direction of the magnetic field. The ratio of the tether's speed to the ion cyclotron frequency which is about 25 m for a LEO is a characteristic length for the phenomena. Whereas for the dimensions of the contact plasma much larger than this value the waves are the conventional Alfvén waves, when the dimensions are comparable or smaller, diffraction effects occur similar to those associated with Fresnal diffraction in optics.

The power required to excite these waves for a given tether current is used to estimate the impedance associated with this mode of charge dissipation. The result, on the order of an ohm, is encouraging.

I. INTRODUCTION

According to Faraday's law of induction, a voltage will be induced in a wire moving across magnetic field lines. Provided there is a stationary return path for current, electrical power can be extracted (an IxB force will act to reduce the relative speed of the wire with respect to the magnetic field), or with the application of a reversed voltage greater than the induced voltage, electrical power will be expended and propulsion will result. In low Earth orbit when the orbital speed, v_c is about $8 \cdot 10^3$ m/s, a 10 kilometer long wire would have an induced voltage of slightly more than 2 kV. Neglecting losses and assuming good electrical contact with the ionosphere, 20 kw of power would be equivalent to a propulsion thrust of about 2.5 N.[1]

The essential elements in the proposed system is the establishment of a good electrical contact with the ionosphere. The most convenient way to establish this contact is with a plasma generator capable of producing plasma in such quantities that it will exceed the ionosphere plasma out to a distance of ten meters or so. Hollow cathodes[2-5] will supply both the plasma need for electrical contact with the ionosphere and electrons needed for the electrical current. Electrical power/propulsion systems using plasma contactors are generally called Plasma Motor/Generators (PMG).

In the literature, most authors[6-10] assume Alfvén waves are responsible for carrying charge from the ends of electrodynamic tethers along magnetic field lines into the ionosphere. The theoretical calculations are extremely long and the discrepancies of their results,

e.g. impedance of the ionosphere vearying from 1 to 10^5 ohms, require that careful study of such calculations be undertaken to identify results common to all calculation and those effected by model dependent assumptions. Consequently, the next section will consider the general solutions of Maxwell's equations under the common assumption that the ionosphere can be considered linear dielectric material. Subsequent sections will consider the effects of tether configuration and frequency bands on the predictions.

II. THE WAVE EQUATION AND ITS SOLUTION

Maxwell's equations for the rext frame of the ionosphere are

$$\vec{\nabla}\cdot\vec{E} = 4\pi\rho \quad , \quad \vec{\nabla}\cdot\vec{B} = 0$$
$$c\,\vec{\nabla}\times\vec{E} = -\dot{\vec{B}} \quad , \quad c\,\vec{\nabla}\times\vec{B} = \dot{\vec{E}} + 4\pi\vec{j} \qquad (1)$$

when $\vec{E}$ and $\vec{B}$ are the electric and magnetic fields, and ρ and $\vec{J}$ are the total charge and current densities.

Using the Fourier transform notation

$$F(\vec{k},\omega) = \left(\frac{1}{2\pi}\right)^2 \int e^{-i(\vec{k}\cdot\vec{x}-\omega t)}\, F(\vec{x},t)\, d^3x\, dt$$
$$F(\vec{x},t) = \left(\frac{1}{2\pi}\right)^2 \int e^{i(\vec{k}\cdot\vec{x}-\omega t)}\, F(\vec{k},\omega)\, d^3k\, d\omega \qquad (2)$$

and combining the latter two Maxwell equations yields the wave equation

$$c^2k^2\,\vec{E} - c^2(\vec{k}\cdot\vec{E})\,\vec{k} = \omega^2\vec{E} + 4\pi i\vec{j} \qquad (3)$$

Treating the ionospheric plasma as medium described by a dielectric tensor $\overleftrightarrow{\varepsilon}$:

$$\vec{D} = \overleftrightarrow{\varepsilon}\cdot\vec{E} \qquad (4)$$

and introducing the current density within the plasma by

$$\vec{j}_p = \overleftrightarrow{\sigma} \cdot \vec{E} = -\frac{i\omega}{4\pi}(\vec{D} - \vec{E}) \tag{5}$$

where $\overleftrightarrow{\sigma} = -\frac{i\omega}{4\pi}(\overleftrightarrow{\varepsilon} - 1)$ is the plasma conductivity tensor, the wave equation can be written as

$$c^2 k^2 \vec{E} - c^2 (\vec{k} \cdot \vec{E}) \vec{k} = \omega^2 \vec{D} + 4\pi i \omega \vec{j}_c \tag{6}$$

where $\vec{j}_c$ is the current density within the conducting tether. Choosing a coordinate system such that $\hat{z}$ is parallel to the geomagnetic field and $\hat{x}$ is the direction of tether motion which is assumed for simplicity to be perpendicular the geomagnetic field, the dialectric tensor of the form[11]

$$\overleftrightarrow{\varepsilon} = \begin{pmatrix} \varepsilon_\perp & +ig & 0 \\ -ig & \varepsilon_\perp & 0 \\ 0 & 0 & \varepsilon_\parallel \end{pmatrix} \tag{7}$$

For frequencies in the region of interest, i.e. much less than the plasma frequency, ω_{pe}, $\varepsilon_\parallel$ is very large and E_z can be neglected compared to D_z in the wave equation.

Without specifying further the frequency dependence of the dialectric tensor or of the tether current density, the wave equation can be solved for the electric field:

$$E_{\substack{x \\ y}} = \left(\frac{4\pi i \omega}{c^2}\right)\left[\frac{\vec{k}\cdot\vec{j}\,(k^2 - k_a^2)k_{\substack{y \\ x}}/k_\perp^2 \mp (\vec{k}\times\vec{j})_z (k_z^2 - k_a^2)/k_\perp^2 \pm i g \omega^2 j_{\substack{y \\ x}}/c^2}{(k^2 - k_a^2)(k_z^2 - k_a^2) - g^2 \omega^4/c^4}\right] \tag{8}$$

where $k_\perp^2 = k_x^2 + k_y^2$ and $k_A^2 = \omega^2 \varepsilon_\perp / c^2$.

As will be explained later, the g^2- term in the denominator is neglible. Consequently the E-field involves Cauchy poles in the

complex k-plane at $k^2 = k_A^2$, $k_z^2 = k_A^2$ and $k_\perp^2 = 0$.

The power needed for the establishment and radiation of electromagnetic fields and the associated charge and current densities is given by Poynting's theorem

$$P = W - \dot{U} \tag{10}$$

where $$W = -\int \vec{j}\cdot\vec{E}\, d^3x$$

$$U = \frac{1}{8\pi}\int (E^2 + B^2) d^3x$$

W is the rate of conversion of mechanical energy to electromagnetic energy and U is the field energy. Since the fields do not change with time in the rest frame of the tether, the time derivative of U will be zero. Further, with the assumption of a cold, collisionless plasma, there is no conversion of electromagnetic energy into mechanical and the power dissipation takes the form

$$P = -\int \vec{j}_c\cdot\vec{E}\, dx^3 \tag{11a}$$

As a result of the tether having a constant speed, v_c, and a constant current, I, both $\vec{j}$ and $\vec{E}$ have Fourier transforms of the form

$$F(\vec{k},\omega) = F(\vec{k})\,\delta(\omega - k_x v_c)$$

which allows the power dissipation to be recast as the real part of a Fourier integral:

$$P = -\frac{Re}{2\pi}\int \vec{j}_c^{\,*}(\vec{k})\cdot\vec{E}(\vec{k})\, d^3k \tag{11b}$$

Since the current density in the tether, $\vec{j}_c$, is proportional to the tether current I, and the electric field is linear in $\vec{j}_c$, the power is proportional to I^2. Consequently the impedance, Z, of the ionosphere required for the maintenance of these charge carrying, electromagnetic waves is given by

$$Z = -\frac{Re}{2\pi I^2} \int \vec{j}_c^{\,*}(\vec{k}) \cdot \vec{E}(\vec{k}) d^3k \qquad (12)$$

The charge density in these waves can be calculated from the continuity equation

$$\dot{\rho}(\vec{x},t) = -\vec{\nabla} \cdot \vec{j}(\vec{x},t) \quad \text{or} \quad \rho(\vec{k},\omega) = +\vec{k} \cdot \vec{j}(\vec{k},\omega)/\omega \qquad (13)$$

where $\vec{j}(\vec{k},\omega)$ is the total current density $\vec{j} = \vec{j}_c + \vec{j}_p = \vec{j}_c - \frac{i\omega}{4\pi}(\vec{D}-\vec{E})$
From the wave equation, it follows that $-\frac{i\omega}{4\pi}\vec{k}\cdot\vec{D} = -\vec{k}\cdot\vec{j}_c$ and thus

$$\rho(\vec{k},\omega) = \frac{c}{4\pi} \vec{k} \cdot \vec{E}(\vec{k},\omega) \qquad (14)$$

As a result of equations (8), (12) and (14) the electric field, the impedance, and the charge density are determined once the frequency dependence of the dialectric tensor and the model used for the current density of tether are specified.

III. MATHEMATICAL ORIGIN OF ALFVÉN WAVES

In the frequency regions were the g^2 - term in the denominator of equation (8) can be neglected the E-field is seen to have Cauchy-poles at $k_z^2 = k_A^2$, $k^2 = k_A^2$ and $k_x^2 + k_y^2 = 0$

where $k_A^2 = \omega^2 \varepsilon_\perp / c^2 = k_x^2 \left(\frac{v_c}{c}\right) \cdot \varepsilon_\perp (k_x v_c)$

Only the first two of these poles involve k_z and thus are candidates for the generation of waves moving along the direction of the

geomagnetic field. As a result, in performing the inverses Fourier transform, the most general k_z-integral will be of the form

$$\int e^{ik_z z}\left[\frac{A}{k_z^2 - k_A^2} + \frac{B}{k^2 - k_A^2}\right] dk_z \qquad (15)$$

where A and B are functions of k which are non-singular in k_z.

The $k_z^2 = k_A^2$ pole is really two poles

$$k_z = \pm k_x \left(\frac{V_c}{x}\right) \varepsilon_\perp \qquad (16)$$

which would result in z-dependencies of the form $exp\left[\pm i k_x \left(\frac{V_c}{c}\right) \varepsilon_\perp^{1/2} z\right]$ When combined with the k_x integration this factor would result in $exp\left[i k_x \left(x \pm \left(\frac{V_c}{c}\right) \varepsilon_\perp^{1/2} z\right)\right]$.

If there is to be radiation into the wake of the tether then two conditions must be satisfied. First, the appropriate pole is the + one of z>o and - om for z<o. The path of integration along the k_z axis should be appropriately chosen to retain only the physical pole. Second, in order that the waves not be damped, i.e. that they are able to carry the charge deep into the ionosphere, requires that $\varepsilon_\perp > o$.
This latter condition will be discussed again in the next section where it will restrict the frequencies leading to wave generation to three regions or bands.

If the above two conditions are satisfied, then each k_x- wave will have a wake angle (see figure 1) given by

$$\alpha(k_x) = \tan^{-1}\left(\frac{V_x}{V_z}\right) = \tan^{-1}\left(\frac{V_c}{V_A}\right) \qquad (17)$$

where $V_A = c/\varepsilon^{\frac{1}{2}}$, is their speed in the z-direction. The k_z poles remaining to be considered come from $k_z^2 = k_A^2 - k_x^2 - k_y^2$. The condition that the waves propagate and carry charge along the z-axis, requires

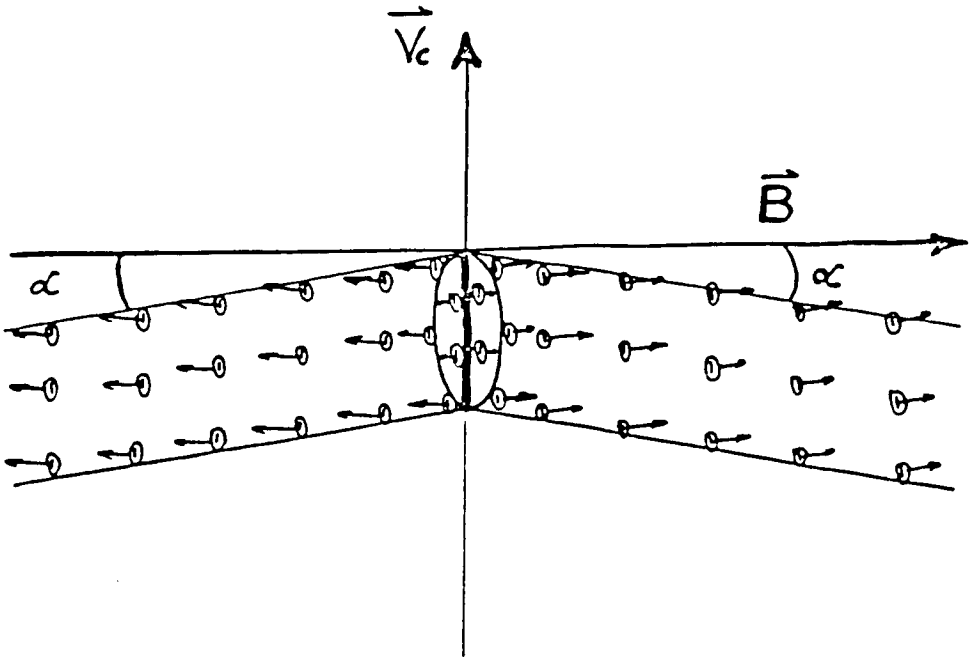


Figure 1. Bottom Tether Wire As Viewed From Earth

that $k_z^2 > 0$, or $k_x^2\left(\left(\frac{v_c}{c}\right)^2\varepsilon_\perp - 1\right) > k_y^2$

Since $(v_c/c)^2 \cong 7\cdot10^{-10}$, only frequency regions where $\varepsilon_\perp > 10^9$ contribute, a severe restriction. Even then, the contributions from the k_x-k_y plane will be very restricted.

Consequenty, only the pole at the $k_z^2 = k_A^2$ will contribute to waves capable of carrying charge into the ionosphere. It should be pointed out that contributions to the integral in equation (12) in which k_z becomes imaginary are themselves imaginary and do not contribute to the impedance.

IV. THE DIELECTRIC TENSOR FOR THE IONOSPHERE

The resonance frequencies for electrons and ions in the ionosphere are due to plasma oscillations and cyclotron motion about the B-field. The dielectric tensor has the form given in equation (7) with

$$\varepsilon_\perp = \tfrac{1}{2}(L+R), \quad g = \tfrac{1}{2}(L-R), \quad \varepsilon_\parallel = 1 - \sum_k \left(\frac{\omega_{pk}}{\omega}\right)^2 \qquad (18a)$$

when

$$R = 1 - \sum_k \left(\frac{\omega_{pk}}{\omega}\right)^2 / (1 + Z_k\Omega_k/\omega)$$

$$L = 1 - \sum_k \left(\frac{\omega_{pk}}{\omega}\right)^2 / (1 - Z_k\Omega_k/\omega)$$

and $\omega_{pk}^2 = 4\pi n e^2/m_k$, $\Omega_k = \frac{eB}{m_k c}$, $Z_k = \begin{cases} + \text{ ions} \\ - \text{ electrons} \end{cases}$

The sums are to be taken over electrons and all ion species. For a two component plasma,

$$\varepsilon_\perp = \frac{(\omega^2 - \omega_L^2)\cdot(\omega^2 - \omega_u^2)}{(\omega^2 - \Omega_i^2)\cdot(\omega^2 - \Omega_e^2)}, \quad g = \frac{\omega\omega_{pe}^2\Omega_e\left(1 - \left(\frac{\Omega_i}{\Omega_e}\right)^2\right)}{(\omega^2 - \Omega_i^2)(\omega^2 - \Omega_e^2)} \qquad (18b)$$

where $2\omega_u^2 = \Omega_i^2 + \Omega_e^2 + \omega_{pi}^2 + \omega_{pe}^2 \mp \sqrt{(\Omega_e^2 - \Omega_i^2 + \omega_{pe}^2 - \omega_{pi}^2)^2 + 4\omega_{pe}^2\omega_{pi}^2}$

are the lower and upper hybrid frequencies. A sketch of $\varepsilon_\perp$ as a

function of ω is shown in figure 2. Clearly the requirement that the waves not be damped, i.e. $\varepsilon_\perp > 0$, is satisfied in only three frequency regions:

Region I $\quad 0 \leq \omega < \Omega_i$ (19)

Region II $\quad \omega_L < \omega < \Omega_e$

Region III $\quad \omega_u < \omega$

According to reference 11, numerical values appropriate to an oxygen-electron plasma are $\Omega_i = 2.0 \cdot 10^6$ Hz, $\omega_L = 3.5 \cdot 10^7$ Hz, $\omega_{pi} = 2 \cdot 10^5$ Hz, $\Omega_e = 5.9 \cdot 10^6$ Hz $\omega_{pe} = 3.4 \cdot 10^7$ Hz, $\omega_u = 3.5 \cdot 10^7$ Hz. In the first region $\varepsilon_\perp \approx \left(\frac{c}{v_{Ao}}\right)^2 / \left(1 - \left(\frac{\omega}{\Omega_i}\right)^2\right)$, where $v_{Ao} = c(\Omega_i/\omega_{pi})$ is the Alfvén velocity, and $g \simeq \frac{\omega}{\Omega_i} \varepsilon_\perp$.

V. MODELS FOR THE TETHER CURRENT DENSITY

The simplest model for the current density of the tether is that of Estes[6]. In this model the vertical tether extends from $-L/2 \leq y \leq L/2$ and is infinitely thin:

$$j_y(\bar{x},t) = I\,\delta(x')\,\delta(z)\,[\theta(y_+) - \theta(y_-)]$$

where $x' = x - v_c t, y_\pm = y \pm L/2$ and the Θ function is given by

$$\Theta(y) = \begin{cases} 1 & y > 0 \\ 0 & y < 0 \end{cases}.$$

There is an infinitely thin horizontal tether at each end of the vertical tether extending from $-L_x/2 \leq x' \leq L_x/2$ such that the currents split at x' = 0 into equal parts, one in the positive x-direction, and another in the negative x-direction:

where $x'_\pm = x' \pm L_x/2 = x - v_c t \pm L_x/2$. Using equation (2) the corresponding Fourier transformed currents are

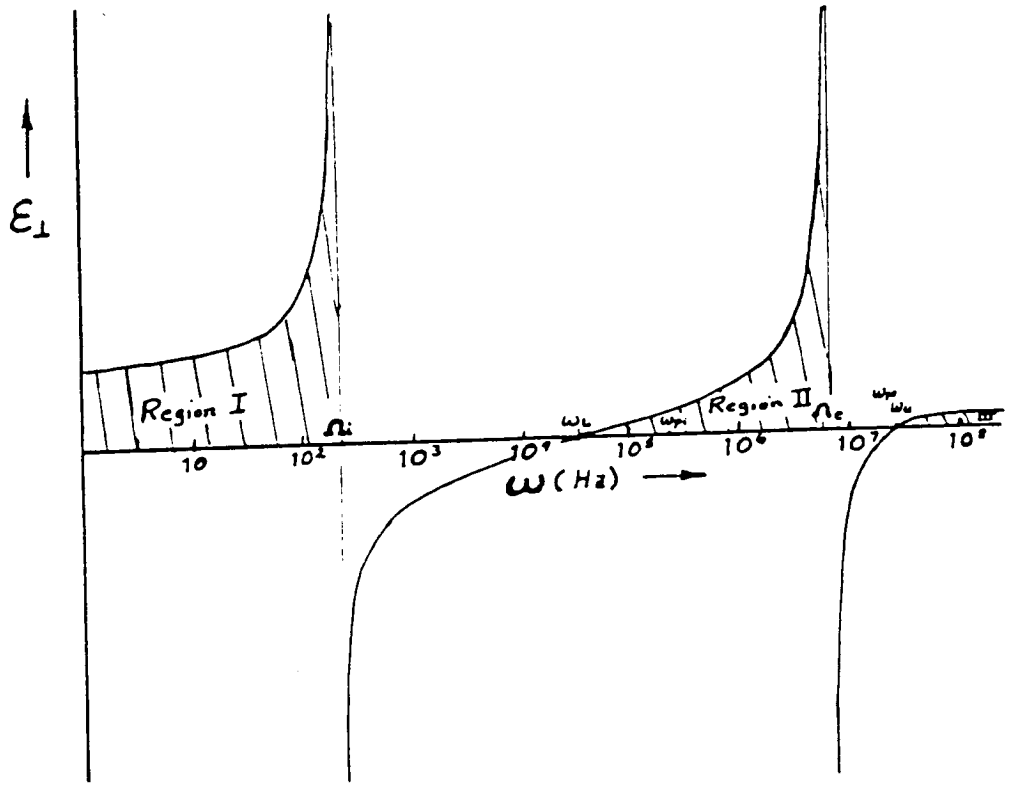


Figure 2. Sketch of $\varepsilon_\perp$ as a Function of Frequency

$$j_x = \frac{I}{\pi}\delta(\omega - k_x v_c)\left[\frac{\sin k_x L_x/2}{k_x L_x/2} - \frac{1}{k_x}\right]\sin k_y L/2 \qquad (20)$$

$$j_y = \frac{I}{\pi}\delta(\omega - k_x v_c)\ \frac{\sin k_y L/2}{k_y}$$

Thus the Fourier transform of $\vec{\nabla}\cdot\vec{j}$ is given by

$$\vec{k}\cdot\vec{j} = \frac{I}{\pi}\delta(\omega - k_x v_c)\ \frac{\sin k_x L_x/2}{k_x L_x/2}\ \sin k_y L/2 \qquad (21a)$$

If, as done in reference (7), the horizontal wires are made into sheets, the result is

$$\vec{k}\cdot\vec{j} = \frac{I}{\pi}\delta(\omega - k_x v_c)\ \frac{\sin k_x L_x/2}{k_x L_x/2}, \sin k_y L/2 \cdot \frac{\sin k_z L_z/2}{k_z L_z/2} \qquad (21b)$$

It is not possible to tell from the above expressions for $\vec{k}\cdot\vec{j}$ whether the vertical wire is infinitely thin or not. Since there is no divergence of $\vec{j}$ coming from the vertical wire, its cross section does not effect $\vec{k}\cdot\vec{j}$. It is of course possible to turn the horizontal sheets into parallelepipeds. To describe the charge carried into the ionosphere by the plasma cloud generated by a hollow cathode, it might suffice to use a spherical or elliptical surface of appropriate size over which the divergence of the current density is constant.

As emphasized by Estes[6] the factor $\left(\frac{\sin k_x L_x/2}{k_x L_x/2}\right)$ in the expressions for $k\cdot j$ acts as a frequency cut-off for finite L_x, i.e. the dominate frequencies are those for which

$$\omega = k_x v_c < \pi v_c/L_x \qquad (22)$$

According to Estes[6], this cut-off for a minimal tether extent ($L_x \cong$ m)

is just below the lower hybrid frequency for the F-layer of the ionosphere. Consequently, in calculating the impedance where the square of this cut-off factor occurs, there will be no significant contributions from any but the lowest frequency band for the PMG project.

VI. RESULTS FOR THE IMPEDANCE AND CHARGE DENSITY

The impedance, Z, of the ionosphere required for the maintenance of the charge carrying electromagnetic waves is given by a k-space integral of $\vec{j}_c^* \cdot \vec{E}$. Using the electric field found in equation (8) yields

$$\vec{j}(\vec{k}) \cdot \vec{E}(\vec{k}) = \frac{4\pi i\omega}{c^2}\left[\frac{|\vec{k}\cdot\vec{j}|^2}{(k_z^2 - k_A^2)k_\perp^2} + \frac{|(\vec{k}\times\vec{j})_z|^2}{(k^2 - k_A^2)k_\perp^2} + O(g^2\omega^4)\right] \qquad (23)$$

As shown in section III, the pole at $k^2 = k_A^2$ gives no significant contribution to the generation of waves and to the impedance. The cut-off factor discussed in the last section, limits the frequencies to those of region I, i.e. $\omega \leq \Omega_i$ and where $g\left(\frac{\omega}{\Omega_i}\right)\varepsilon_\perp$ Consequently, using the simplest model given by equation (21a) yields

$$Z = -\frac{Re}{\pi^2 c^2/2}\int_{-\kappa}^{\kappa} dk_x \int_{-\infty}^{\infty} dk_y \int_{-\infty}^{\infty} dk_z \frac{ik_x v_c \left(\frac{\sin k_x L_x/2}{k_x L_x/2}\right)^2 \sin^2 k_y L/2}{(k_z^2 - k_A^2)k_\perp^2} \qquad (24)$$

where $k_A = k_x\left(\frac{v_c}{v_A}\right)\cdot\left(1 - \left(\frac{k_x v_c}{\Omega_i}\right)^2\right)^{-\frac{1}{2}}$ and $\kappa = \frac{\Omega_i}{v_c}$. With the appropriate contour for the k_z - integration, the expression for the impedance becomes

$$Z = \frac{Re}{\pi c^2/2 v_m}\int_{-\kappa}^{\kappa} dk_x \int_{-\infty}^{\infty} dk_y \left(1 - \left(\frac{k_x v_c}{c}\right)^2\right)^{1/2} \frac{\left(\frac{\sin k_x L_x/2}{k_x L_x/2}\right)^2 \sin^2 k_y L/2}{k_x^2 + k_y^2}$$

The integration over dk_y is easily done to give

$$Z = 2\frac{V_A}{c^2}\int_0^1 \left(\frac{\sin k\bar{L}_x/2}{k\bar{L}_x/2}\right)^2 (1-e^{-k\bar{L}})\sqrt{1-k^2}\,\frac{dk}{k} \qquad (25)$$

where $\bar{L} = L/d$, $\bar{L}_x = L_x/d$ with $d = v_c/\Omega_i \cong 25$ meters.

The difference of this final expression for Z and that given in reference (6) is just the contribution for $j_x E_x$ which that author neglected in his determination of Z.

If the limit of $L_x << d << L$ is taken, then

$$Z \to 2\frac{V_A}{c^2}(\ln 2L + \gamma - 1) \qquad (26)$$

where γ is the Euler constant. In this limit the horizontal extent of tether is no longer important and agreement is obtained with reference 7, for their contribution from the first frequency region. Clearly in this limit, the physically important frequency cut-off factor coming from the finite horizontal extend of the tether system is turned off. Consequently, frequencies are no longer limited to the first band and the higher bands could give the significant contributions found by reference 7 for a horizontal extend of a few tenths of a millimeter, i.e. $Z_I \cong .35\Omega$, $Z_{II} \cong 10^5\Omega$ and $Z_{III} \cong 10^4\Omega$. Such extents are unphysical and one should use the results of this reference best as a mathematical check.

A second limit in which $L_x >> L >> d$ is perhaps more physical. In this limit

$$Z \to \frac{2\pi V_A}{c^2}\left(\frac{L}{L_x}\right) \qquad (27)$$

first obtained by reference 10. However this limit could not describe the PMG project since a plasma having such a large extend in the direction of motion would similarly extend in the vertical direction and envelope both ends of the tether and thus short-circuit the system.

The inclusion of the factor $\frac{\sin k_z L_z/2}{k_z L_z/2}$ of equation (21b), due to the extent in the z-direction, brings in a factor of

$\left(\frac{\sin k_A L_z/2}{k_A L_z/2}\right)^2$ into the impedance calculation. Since $k_A = k_x \left(\frac{v_c}{c}\right)\varepsilon_\perp^{\frac{1}{2}}$ this factor will further strengthen the frequency cut-off.

The plasma generated by a hollow cathode will itself be distorted by the geomagnetic field. Due to the large value of the conductivity in the direction of the geomagnetic field compared to that perpendicular to the field, one could expect the effective extend of the plasma cloud in the field direction, L_z, to be very much larger than that in the perpendicular direction, L_x. It might well be that in the PMG project the spatial extend of the plasma cloud along the field lines will be the critical parameter.

Turning to the predictions of the models for charge density in the Alfvén wings, equations (8) and (14) yield

$$\rho(\vec{k},\omega) = \frac{\omega}{c^2}\left[\frac{k\cdot j}{k_z^2 - k_A^2} + \frac{i(k\times j)_z q\omega^2/c^2}{(k^2-k_A^2)(k_z^2-k_A^2)} + O(q^2\omega^4/c^4)\right] \quad (28)$$

Rewriting the denominator of the second term as

$$\frac{1}{(k^2-k_A^2)(k_z^2-k_A^2)} = \left[\frac{1}{k_z^2-k_A^2} - \frac{1}{k^2-k_A^2}\right]\frac{1}{k_\perp^2}$$

the pole structure remaining, after the non-contributiong pole at $k^2 = k_A$ is discarded, is just that encountered in equation (24). The result of the k_z and k_y integration is a contribution similar to that resulting from the first term in the equation (28) but reduced by a factor of $(v_c/c)^2$.

The contribution of the first term gives

$$\rho(\vec{x},t) = \frac{I v_A}{2\pi c^2}\left(\delta(y-\tfrac{L_y}{2}) - \delta(y+\tfrac{L_y}{2})\right)\int_0^1 dk \left(\frac{\sin k\bar{L}_x/2}{k\bar{L}_x/2}\right)\cos k(\bar{x}+|\bar{z}'|)\cdot\sqrt{1-k^2}$$

where $\bar{x} = x/d$, $\bar{z}' = \bar{z}\,(v_c/v_A)\,(1-k^2)^{-\frac{1}{2}}$.

Only for $L_x >> d$ does the upper limit of the integration become unimportant and diffraction effects due to the $(1-k^2)^{\frac{1}{2}}$ factor relating z' and z result in the normal sharp Alfvén wings. Figure 3 shows the situation for the case of L_x = 10 m and L_x = 800m.

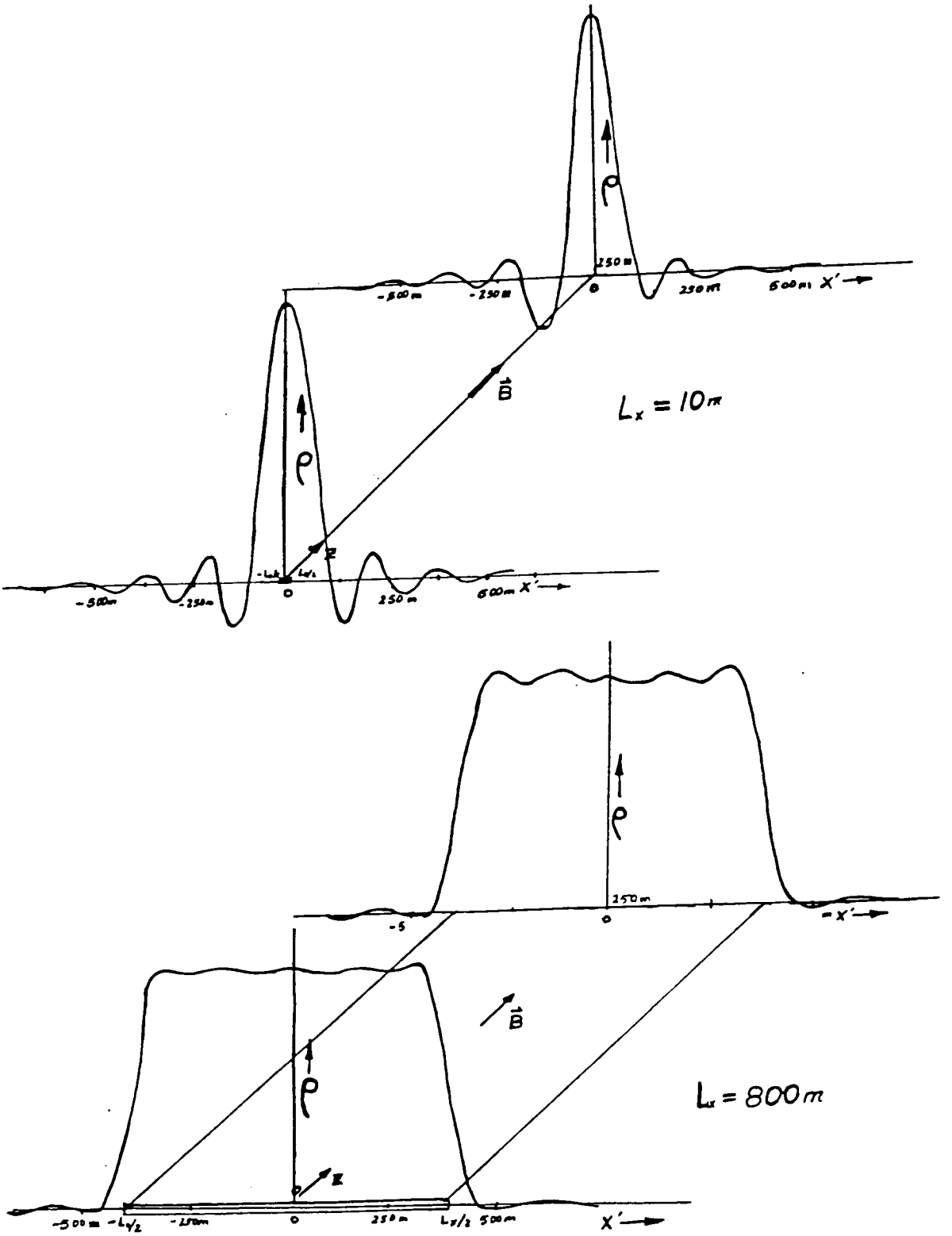


Figure 3. Charge Density in Upper Wing for L_x = 10 m and 800 m.

VII. CONCLUSION

This report has attempted to clearly point out the assumptions made and common results found in theoretical models put forth to explain the dissipation of charge into the ionosphere from the ends of a conductive tether. The most dubious of the assumption is that of treating the ionosphere in a "cold plasma" dialectric medium. Once the assumption is made the spatial extend of the tether need be modeled in a reasonable manner. The dimensions of the region over which the charge is released into the ionosphere, is shown to play a crucial role in restricting the phenomena to low frequencies. If a physically reasonable size is used, calculations show that the models predict an impedance on the order of a few ohms.

The adaption of this model to describe the PMG project is not easy since the plasma clouds generated at the ends of the tether will themselves be strongly affected by the geomagnetic field. Still it seems reasonable to accept the impedance value mentioned above as a good first approximation.

REFERENCES

1. Hite, G. E., and J. E. McCoy, "Electrodynamic Tether", accepted for publication in the American Journal of Physics (1987).

   McCoy, J. E., "Plasma Motor-Generator System Designs for Power and Propulsion Interaction", Conference on Tethers in Space, September 1986.

   McCoy, J.E., "Electrodynamic Tethers, I. Power Generation in LEO, II. Thrust for Propulsion & Power Storage", Intern, Astronautical Federation, 35 Congress, Paper 440, (Oct. 1984).

2. Krishnan, M., R. G. John, W. F. von Jaskowsky, and K. E. Clark, "Physical Processes in Hollow Cathodes", AIAA Journal Vol 15 No. 9, pp 1217-1225 (1977).

3. Siegfried, D. E., and P. J. Wilbur, "Studies on an Experimental Quartz Tube Hollow Cathode, Electric Propulsion and its Applications to Space Missions", Vol 79 of Progress in Astronautics and Aeronautics, pp 262-277 (1981).

4. Siegfried, D. E., and P. J. Wilbur, "An Investigation of Mercury Hollow Cathode Phenomena", 13th Intern. Electric Propulsion Conf., (April 1978).

5. Siegfried, D. E., " A Phenomenological Model for Orified Hollow Cathodes", NASA CR 168026, (Dec. 1982).

6. Estes, R. D., "Alfvén Waves from an Electrodynamic Tethered Satellite System", accepted for publication in J. of Geophy. Res. (1987).

7. Barnett, A. and S. Olbert, "Radiation of Plasma Waves by a Conducting Body Moving through a Magnetized Plasma", J. of Geophys. Res. 91, 10, 117 (10.135)(1986).

8. Dobrowolny, M. and P. Veltri, "MHD Power Radiation by a Large Conductor in Motion Through a Magnetoplasma", IL Nuovo Aminto Vol 9C No. 1, 27-38 (1986).

9. Rasmussen, C. E., P. M. Banks, and K. J. Harker, "The Exitation of Plasma Waves by a Current Source Moving in a Magnetized Plasma: The MHD Approximation", J. of Geophys. Res. 90 No. A1, 505-515 (1985).

10. Drell, S. D., H. M. Foley, and M. A. Ruderman, "Drag and Propulsion of Large Satellites in the Ionosphere: An Alfvén Propulsion Engine in Space", J. of Geophys. Res. 70 No. 13. 3131-3145 (1965).

11. Bauer, S. J., Physics and Chemistry in Space, 6 Physics of Planatary Ionosphere, Springer-Verlag p. 135 (1973).

HIGH DATA RATE MODEM SIMULATION FOR THE SPACE STATION MULTIPLE-ACCESS COMMUNICATIONS SYSTEM

Final Report

NASA/ASEE Summer Faculty Fellowship Program -- 1987

Johnson Space Center

Prepared by:	Stephen Horan, Ph.D.
Academic Rank:	Assistant Professor
University & Department:	New Mexico State University Department of Electrical and Computer Engineering Las Cruces, New Mexico 88003

NASA/JSC

Directorate:	Engineering
Division:	Tracking and Communications
Branch:	Systems Analysis Office
JSC Colleague:	James E. Ratliff
Date:	August 7, 1987
Contract Number:	NGT-44-001-800

ABSTRACT

The communications system for the Space Station will require a space-based multiple-access component to provide communications between the space-based program elements and the station. A study was undertaken to investigate two of the concerns of this multiple-access system, namely, the issues related to the frequency spectrum utilization and the possibilities for higher-order (than QPSK) modulation schemes for use in possible modulators and demodulators (modems).

As a result of the investigation, may key questions about the frequency spectrum utilization were raised. At this point, frequency spectrum utilization is seen as an area requiring further work.

Simulations were conducted using a computer-aided communications system design package to provide a straw-man modem structure to be used for both QPSK and 8-PSK channels. Areas of further work on the modem design are identified.

INTRODUCTION

The Space Station (SS) will have a Multiple-Access Communications System (MACS) as one of its subsystem elements. The general structure and requirements of the MACS have been described in [1]. The MACS will be responsible for the space-to-space segment communications as part of the overall SS communications system. The MACS will need to interface with the on-station communications networks and the space-to-ground communications network.

The following will all be users of the MACS at one time or another:

a) NSTS (space shuttle),
b) Orbital Transfer Vehicle (OTV),
c) Orbital Maneuvering Vehicle (OMV),
d) Flight Telerobotic Servicer (FTS),
e) Extra-Vehicular Activity/Extravehicular Mobility Unit (EVA/EMU),
f) Co-Orbiting Platform (COP),
g) Free-Flyer (FF),
h) Mobile Servicing Centre (MSC)

Each of these users will have different data requirements by data type and data rate. The types and rates which have been identified for use on the MACS are as follows:

a) audio, command (CMD), telemetry (TLM) and user data channels each at 128 Kbps,
b) video channels at 22-25 Mbps,
c) heads-up displays (HUD) at 400 Kbps
d) emergency safety link (ESL) at a rate to be determined but most likely with an upper limit of 128 Kbps.

A matrix showing the SS elements with their associated data types and rates is given in Table 1.

PROJECT OBJECTIVES

Of the broad possible areas of MACS design, it was decided to concentrate on the possibilities for using higher-order modulation schemes for digital data transmission and the associated frequency planning to be used with such a scheme.

In the area of frequency planning, an investigation of the required bandwidth for the data channels was to be performed and a possible strategy for the MACS was to be developed. Also, areas of concern relating to the frequency plan were to be identified.

For the higher-order modulation study, a candidate modem structure which could be used in the SS environment and simulated via the available computer-aided design software was to be provided. This candidate modem would then be used as the basis for more detailed further study.

FREQUENCY PLAN

In order to plan the spectrum, an assessment of the total data needs from all users is needed. Next, a survey of the available choices needs to be made and then identify candidate placements for the MACS frequency spectrum.

FREQUENCY SPECTRUM UTILIZATION

To develop the frequency plan, an accounting of the the data-link requirements is made from the data presented in Table 1. Based upon these estimates, the forward and return

Table 1. Data requirements for Space Station Multiple Access Communications System users.

User	# Links	Max Range	Forward	Type	Return	Type
FF	0(8)	2000 Km	CMD	NB	TLM	NB
			Data	NB	Data	NB
					Video	WB
COP	0(1)	2000	CMD	NB	TLM	NB
					Video	WB
NSTS	1(2)	37	Audio	NB	Audio	NB
			Data	NB	Data	NB
OMV	1(2)	37(185)	CMD	NB	TLM	NB
					Video	WB
EMU/ EVA	2(4)	1	Audio	NB	Audio	NB
			HUD	MB	TLM	NB
					Video	WB
MSC	1(1)	0.1	CMD	NB	TLM	NB
					(3)Video	WB
FTS	1(1)	0.1	CMD	NB	TLM	NB
			ESL	NB	(4)Video	WB
OTV	0(1)	185	CMD	NB	TLM	NB
					Video	WB

Notes: The number of links shows the initial configuration with the possible growth in parenthesis; the link type of NB corresponds to 128 Kbps, MB to 400 Kbps, and WB to 25 Mbps. Data taken from [2] and [3].

link requirements for Initial Operating Conditions (IOC) and projected growth (Growth) are summarized as follows:

a) forward links will require for IOC(Growth)
   1) 8(31) 128-Kbps links,
   2) 2(4) 400-Kbps links,
   3) no forward video links (at present);
b) return links will require for IOC(Growth)
   1) 9(34) 128-Kbps links,
   2) no requirement fpr 400-Kbps links,
   3) 10(16) 25-Mbps links.

There is no requirement for 100% availability on all links simultaneously. This leads to a suggested frequency plan as shown in Figure 1 where a design limitation of 300 MHz total bandwidth was used. This plan assumes that there will be some form of time-division access to the wide-band video channels. This plan will also allow for all of the medium-band and narrow-band channels, i.e., 400-Kbps and 128-Kbps, respectively, to be accomodated. In planning this spectrum utilization an assumption was made that a spectrally-efficient form of modulation would be found so that the up-to 100-Mbps data sources would occupy the same spectral bandwidth as a 25-Mbps source would.

FREQUENCY-BAND SELECTION

Based on [3], Ku-Band (12-18 GHz) was used as the frequency band for placement of the MACS. However, Ku-Band is presently heavily allocated by international agreement as shown in [5]. Another possible frequency band to be used for the MACS is Ka-Band (27-40 GHz). This spectral band is presently not heavily allocated. However, there may be problems related to directivity of antennas among other areas for this to be a realistic contender at this time.

MODEM DESIGN

The second phase of this project was to consider candidate modem structures for higher-order modulation applications.

While Quadrature Phase Shift Keying (QPSK) is considerd to be quite standard in many current applications, the required video data links push the design into the area of greater spectral efficiency. This is then the motiviation behind looking at higher modulation orders. In this section, the design drivers for a MACS modem and the modem structures investigated during this project are presented.

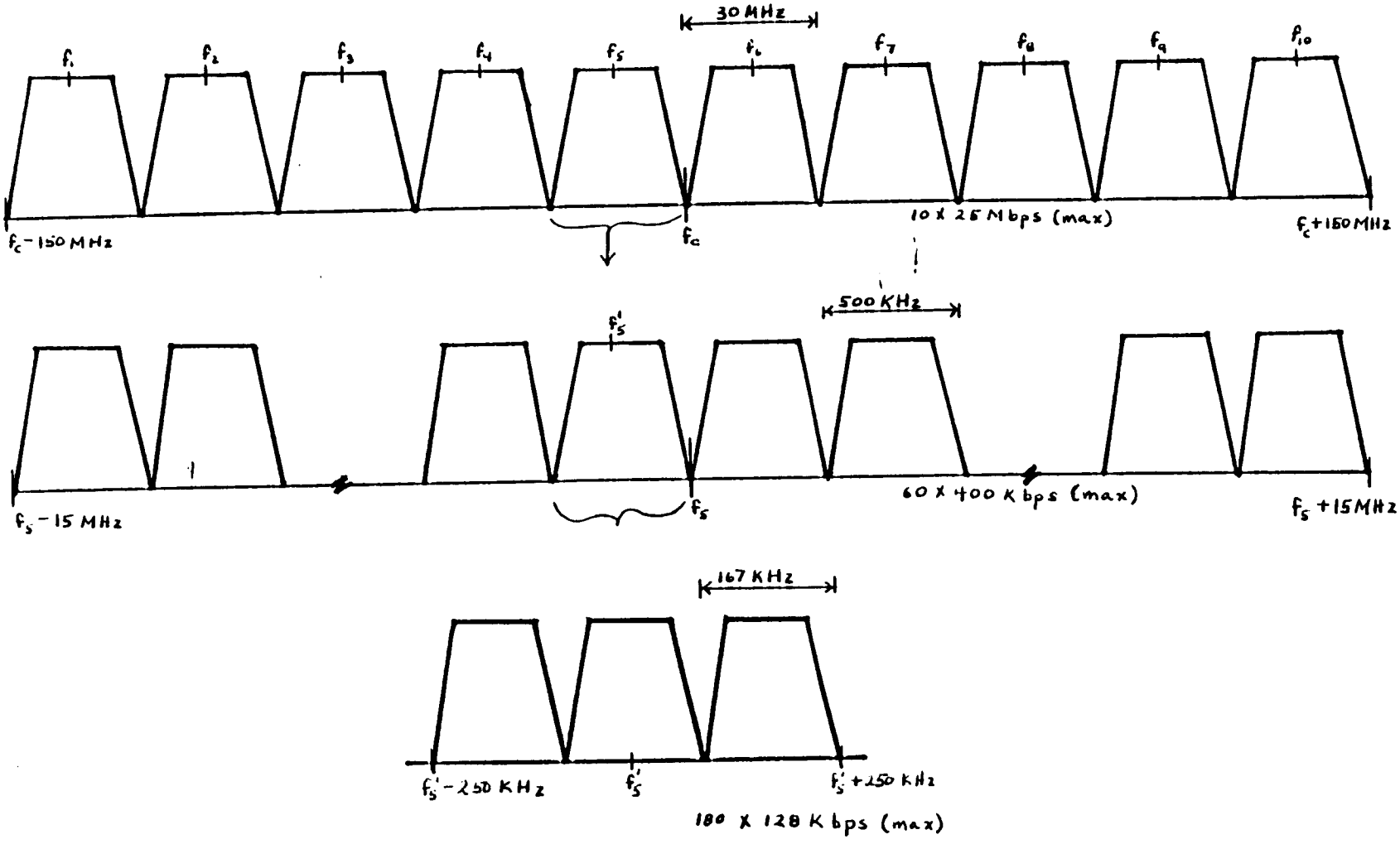


Figure 1. Candidate spectrum utilization for forward and return links. Assumes a spectral efficienoy of 83.3% would be used in the modem realization.

DESIGN DRIVERS

The design drivers on the selection of modem for the MACS are

a) a basic compatability with current technology which would allow existing QPSK sources to make some use of the MACS,
b) the design should allow for an evolution from QPSK to be made in a controlled growth environment,
c) the theoretical Bit Error Rate (BER) performance should not intrinsically require extensive data or symbol coding to achieve a design specification of 10**-5,
d) the modem package should not tend towards requiring extensive amounts of signal processing equipment to do its job, i.e., be portable and battery operated.

With these ground rules, selection of a higher-order modulation scheme and candidate modulators and demodulators proceeded.

SELECTION OF MODULATION FORMAT

There are two basic modulation schemes which can reasonably be developed based upon a QPSK structure: some form of higher-order PSK modulation or Quadrature Amplitude Modulation (QAM). This is because mathematically, QPSK can be seen as being either a case of M-ary PSK or M-ary QAM. This then rules out some form of Frequency Shift Keying (FSK) being considered.

It can be shown [6] that the M-ary PSK and M-ary QAM both have theoretical spectral efficiencies of ℓ /2 bits/sec/Hz where

$$\ell = \log_2(M).$$

Therefore, there is no intrinsic preference for an overall modulation scheme based upon theoretical efficiency arguments.

The required energy-per-bit to noise-power ratio (Eb/No) to achieve a specified BER performance for the different M-ary PSK and QAM modulation formats is illustrated in [7]. Theoretically, QPSK requires an Eb/No of 13 dB at the specified BER and one would like to stay as close to this as possible. The next higher orders of PSK and QAM above QPSK are 8-PSK and 16-QAM which require an Eb/No of 18 dB and 20 dB, respectively. Above that, it would appear that we would require significant amounts of coding to close the link and subsequently cut the data throughput so the search for an

appropriate higher modulation scheme stops here. At this level, the basic goal of providing enough video channels for the MACS can be met with a combination of frequency- and time-division multiplexing. Since 8-PSK modulation would appear to require less power to close the link, it was decided to concentrate on it for this study.

CANDIDATE MODULATORS

It has been shown [8] that M-ary PSK modulators can be built by cascading stages of QPSK modulation. If $\log_2(M)$ is not evenly divisible by four, then some form of data mapping must be provided to synthesize the full signal constellation. A candidate modulator to do just this was proposed to NASA/JSC [9]. This modulator had the additional feature of allowing switching so that it could be run as either a standard QPSK modulator or as an 8-PSK modulator. As originally presented, this modulator had the basic QPSK channels and the added channels reversed. The mapping was modified to correct this condition. The resulting signal constellation and mapping function are shown in Figure 2.

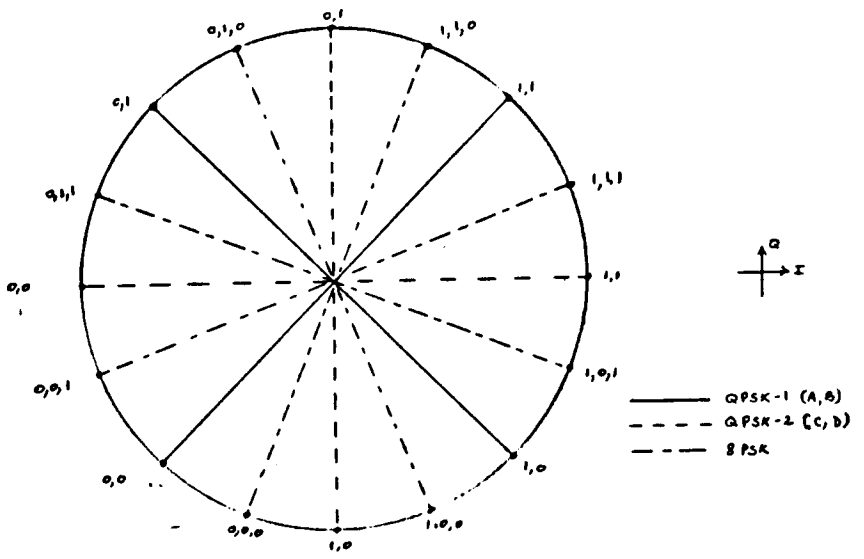
CANDIDATE DEMODULATORS

A possible demodulator for the 4/8-PSK modulator described above was also provided to NASA/JSC [9]. This demodulator was basically the inverse process of the modulator and contains two phase-staggered QPSK demodulators. While this structure shows every reasonable probability of working, it was decided to investigate the possibility of making this structure simpler and possibly even allowing one coherent demodulator work for both QPSK and 8-PSK. A literature search was performed to identify candidate demodulators and they are as follows:

a) Polarity Costas Loop [10]
b) combined Squaring/Quadrupling Loop [11]
c) Tanlock Loop [12].

The Polarity Costas Loop (PCL) structure is an extension of the typical Costas Loop structure for demodulating QPSK signals but with the hard limiters replaced by multi-level limiters. The resulting demodulator is only slightly more complicated than a normal QPSK demodulator. However, the PCL will require a tight Automatic Gain Control (AGC) to work properly. Since the MACS will probably require an AGC this is not seen as being too stringent of a drawback.

The combination of a squaring/quadrupling loop, while similar to QPSK demodulators, would essentially require two demodulators in one package to have switchable 4- and 8-PSK.



(a)

Input Data			Modulator Input				Data Recovery Logic d1	d2	d3
d1	d2	d3	A	B	C	D	sgn(I)	sgn(Q)	|I|>|Q|
0	0	0	0	0	1	0	0	0	0
0	0	1	0	0	0	0	0	0	1
0	1	0	0	1	0	1	0	1	0
0	1	1	0	1	0	0	0	1	1
1	0	0	1	0	1	0	1	0	0
1	0	1	1	0	1	1	1	0	1
1	1	0	1	1	0	1	1	1	0
1	1	1	1	1	1	1	1	1	1

to QPSK #2
to QPSK #1
Added data Channel
Permanent data Channels

(b)

Figure 2. The 8-PSK signal constellation and data mapping. (a) The two QPSK signal constellations and the resulting 8-PSK constellation. (b) The three-to-four-bit data input and output mapping (octal mapping).

The tanlock loop, while being a very efficient digital phase-locked loop for tracking M-ary PSK signals, may have some implementation problems because arctangent functions would need to be synthesized at high data rates. However, the tanlock loop does offer more immunity to input signal gain variations than does the PCL.

MODEM STRAW-MAN DESIGN

Based upon the available choices at this time for a modem structure the following components were chosen upon which to perform simulations:

a) the 4/8-PSK modulator with signal constellation correction,
b) the Polarity Costas Loop demodulator for in-phase (I) and quadrature-phase (Q) channel extraction,
c) a data extractor which recovers the data channels from the I and Q outputs of the PCL and has a data synchronizer with data clock recovery.

The PCL was chosen because of its ability to extract both QPSK and 8-PSK from a signal. The data extractor performs the decoding logic and then is followed by a data clock extractor and synchronizer. The data clock extractor was developed by TRW and used in the simulator without modification. The data synchronizer was a set of sample-and-hold circuits synchronized to the data clock extractor. The block diagrams for these components will be given below when the simulations are described in more detail.

SIMULATION RESULTS

Basic simulations of the candidate modem structure were performed during the project period using the Computer-Aided Design (CAD) system available within the Tracking and Communications Division of JSC. In this section we will describe the simulation environment and the models used in the simulation runs.

SIMULATION ENVIRONMENT

The CAD system used to perform the simulations is a commercially-available product known as the Block Oriented System Simulator (BOSS) [13] which runs on a Digital Equipment VAX Station II workstation. For these simulations, version 1.1 of BOSS was used. BOSS is a menu-driven package using both a mouse and keyboard entries to develop and document system modules. The philosophy of BOSS is to develop standard modules at the block-diagram level and then used these

modules to construct larger modules until an end-to-end system is designed. BOSS provides a variety of basic blocks. If a more complicated module is needed then these basic blocks along with any other blocks the user has created may be combined to build new blocks. Eventually, the blocks are combined to form the end-to-end system.

Simulations of the system proceed by having the user specify simulation duration, time step, and any necessary parameters to make the modules work properly. Usually, the module parameters specified at simulation time are those design parameters, e.g. filter bandwidth, which one is trying to optimize for the system. Prior to starting the simulation, "probes" on signal paths may be specified for data collection. Families of simulations may then be run to optimize parameters and to generate signal plots for documentation and anlaysis.

MODEM BLOCK DIAGRAMS

The candidate test system consisted of a 8-PSK transmitter module and the combination PCL and data recovery modules for a receiver. The combined system is shown in Figure 3 while the modulator and demodulator subsystems are shown in Figure 4. At this point, no system noise sources were introduced. The 8-PSK transmitter consisted of a three-channel random data source (with each channel seeded with different starting numbers), the candidate 4/8-PSK modulator (set for 8-PSK only), and an output bandpass filter. The block diagram for the 4/8-PSK modulator is illustrated in Figure 5. The output bandpass filter was set for first-order Butterworth characteristics and a single-sided bandwidth of five times the data rate which effectively gave an "infinite-bandwidth" system.

The receiver consisted of the PCL demodulator followed by a data recovery module. The block diagrams for these modules are illustrated in Figure 6. Internal to the data recovery module is the TRW-developed data transition tracking loop which was used to produce data outputs at constant relative phase.

SIMULATION RUNS

During the time of this project, only one type of simulation was able to be performed. This simulation passed 25 bits through the test system to verify correct operations of the components on the simulator. Although the system was set for 8-PSK only, the correct operation of the demodulator shows that it can also be used to demodulate QPSK as well.

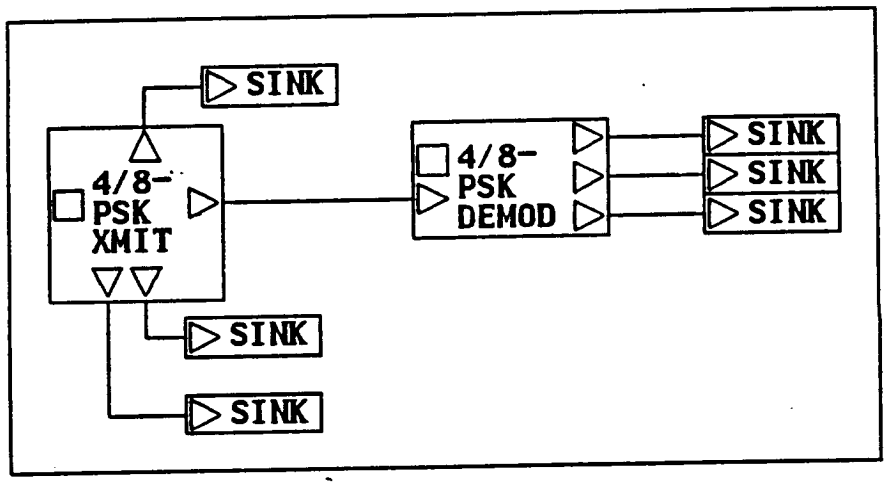
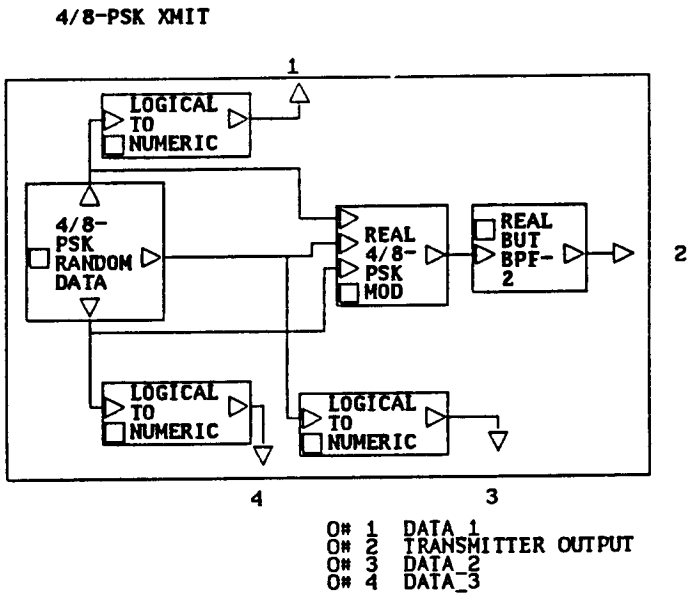
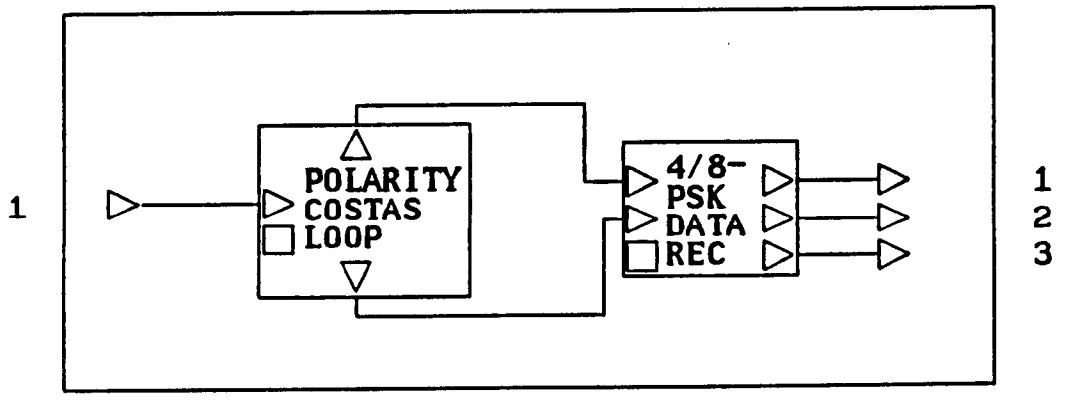


Figure 3. The 4/8-PSK test system block diagram used for the BOSS simulations.



(a)



I# 1 CARRIER_INPUT

O# 1 RECOVERED_DATA_CHANNEL_1
O# 2 RECOVERED_DATA_CHANNEL_2
O# 3 RECOVERED_DATA_CHANNEL_3

(b)

Figure 4. The expanded modulator and demodulator block diagram used for the BOSS simulations. (a) the modulator. (b) the demodulator.

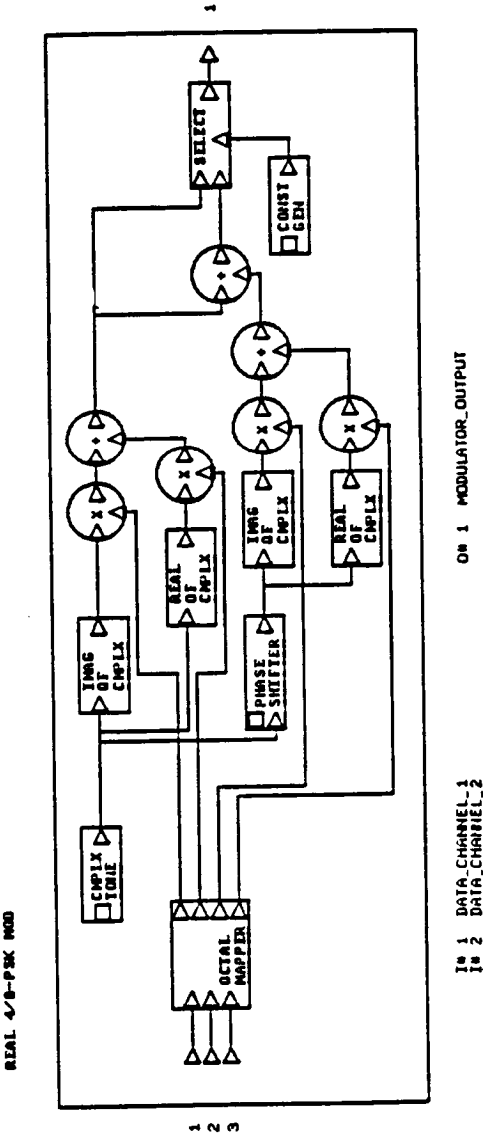
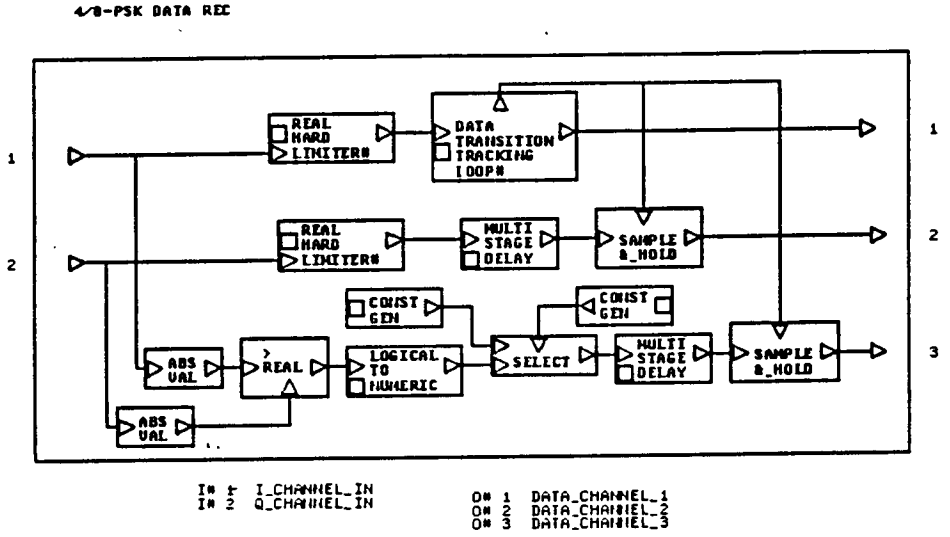
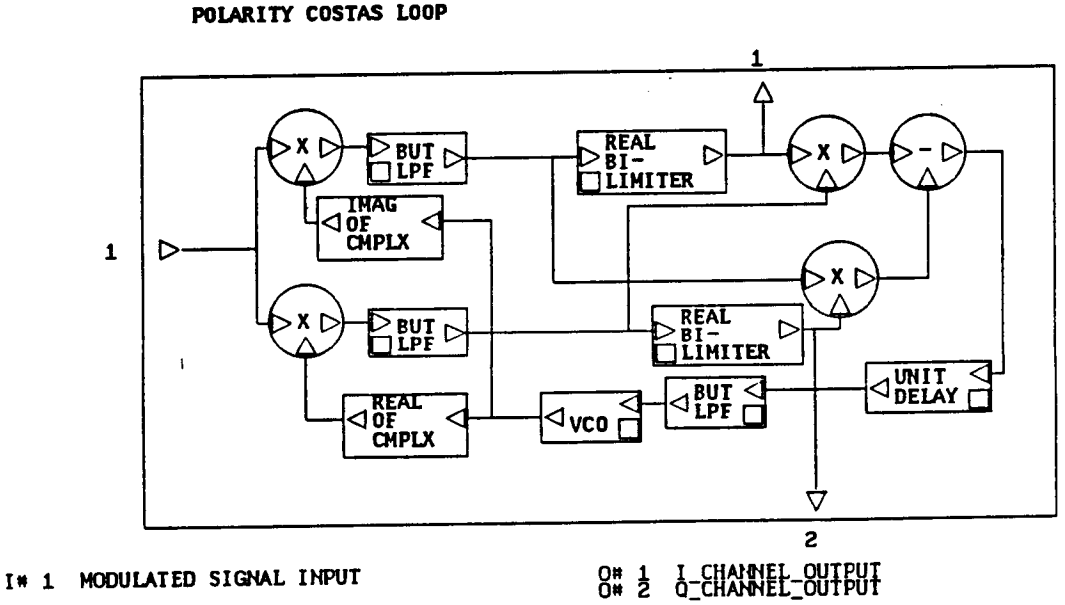


Figure 5. The expanded block diagram for the 4/8-PSK modulator used for the BOSS simulations.



(a)



(b)

Figure 6. The BOSS simulation receiver structure. (a) the Polarity Costas Loop structure, (b) the 4/8-PSK Data Recovery Module.

FINDINGS

In this section, we will summarize the major findings of this project in the areas of frequency spectrum usage and simulation of the candidate modem.

FREQUENCY UTILIZATION

Based upon the brief look at the frequency planning issue, no obvious choice for a location in Ku-Band was found to place the MACS. What was found is that there are areas which still need to be worked on to give a better understanding of the design issues. These issues include the following points:

a) what will be the optimum spectral placement of the MACS based upon
   1) non-interference with Tracking and Data Relay Satellite System (TDRSS) links used by SS and others,
   2) non-interference with allocated satellite services,
   3) link power budget and expected interference with and from ground sources;
b) will the spectrum need to be broken onto forward and return bands and if so what will need to be their relative widths;
c) are there any spectral shaping constraints, e.g., from the FCC, which must be adhered to and which may affect the design;
d) how many intermidiate frequencies and what type of signal gain control will be needed in the transmitters to multiplex the low and high data rates;
e) might a different frequency band be better, e.g., Ka-Band or might a split be preferable where the near users requiring low directivity use Ku-Band and the far users employ the intrinsic directivity of Ka-Band.

SIMULATION RESULTS

In the second area, the baseline infinite-bandwidth candidate modem system with no noise was able to be constructed to verify the correct operation of all of the sub-components. There are further studies which need to be performed to extend the design. The types of simulations which need to be performed are as follows:

a) linear channel simulations to assess the effects of finite bandwidth and Gaussian noise on the modem performance;
b) simulations to assess effects of non-linearities in

the expected channel due to
1) amplifier performance,
2) multipath and fading in the channel,
3) adjacent and co-channel interference;

c) the effects of AGC performance on the demodulation performance of the PCL.

These simulations should also be run with other demodulation techniques not studied here to determine the most robust design.

There is also a need for a study to determine the optimum method for running a detailed simulation. Rough estimates of the required computer time to run a detailed simulation indicate substantial fractions of a year may be necessary. There exist techniques [14] to cut this down by orders of magnitude in linear systems. How these techniques may be incorporated into BOSS and how they may be extended to non-linear channels needs to be investigated.

CONCLUSIONS

The overall conclusions of this study are as follows:

a) the issue of the frequency plan for the MACS is not yet settled, however, needs to be done soon because it will affect the design of the modems used and affect the way in which the MACS is managed;
b) the 4/8-PSK modulator with associated PCL and data recovery functions can make an initial modem structure for further study; there are many design issues to be settled but it represents a reasonable start because the combination holds out the promise of increased spectral efficiency over QPSK, it can be an evolutionary design, and the same demodulator can be used for both QPSK and 8-PSK.

As pointed out in the Findings above, both areas have further issues to be studied.

REFERENCES

[1] S.B. Franklin, C.J. Langton, R.E. Vaughan, and R.M. Ward, "End-to-end Communications for Space Station," Conference Record 1986 Global Telecommunications Conference, pp. 832-837 and

N.A. Olsen, "Space Station Multiple Access Communica-

tions System," ibid., pp. 843-846.

[2] Rockwell International, "Communications and Tracking Report," Space Station Work Package 2, NAS9-17365, DR-02, Vol. 12, Dec. 19, 1986.

[3] Goddard Space Flight Center, "Baseline Configuration Document," III.T. Fligt Telerobotic Servicer, SS-GSFC-0010, Jan. 12, 1987 and

_____, "Space Station Flight Telerobotic Servicer Contract End Item Specification," SS-GSFC-0019, Dec. 15, 1986.

[4] Space Station Program Office, "Architectural Control Document Communications and Tracking System," JSC-30260, undated.

[5] T.T. Ha, Digital Satellite Communications, (New York: Macmillan, 1986) pp. 4-5.

[6] L.W. Couch, Digital and Analog Communication Systems, 2nd ed. (New York: Macmillan, 1987) pp. 323-325.

[7] K. Feher et al., Telecommunications Measurements, Analysis, and Instrumentation, (Englewood Cliffs, NJ: Prentice-Hall, 1987) p. 44.

[8] W.C. Lindsey and M.K. Simon, "Carrier Synchronization and Detection of Polyphase Signals," IEEE Trans. Commun., vol. COM-20, pp. 441-454, June 1972.

[9] S. Udalov and J. Dodds, "Design Considerations for a Universal Multi-Channel Modem for Space Station Ku-Band Links," NAS9-17414, Axiomatix Report No. R8610-1, Feb. 28, 1987.

[10] H.C. Osoborne, "A Generalized 'Polarity-Type' Costas Loop for Tracking MPSK Signals," IEEE Trans. Commun., vol. COM-30, pp. 2289-2296, Oct. 1982.

[11] C.J. Wolejsza, D. Chakraborty, and V. Uzunoglu, "A Universal Modem Design for Digital Satellite Communications," Conference Record 1976 International Communications Conference, pp. 3-12 - 3-18.

[12] H.J. Kim, C.K. Un and J.C. Lee, "The N-Phase Digital Tanlock Loop for Tracking Supressed-Carrier N-ary PSK Signals," IEEE Trans. Commun., vol. COM-33, pp. 904-910, Sept. 1985.

[13] H.T. Mouftah and K.S. Shanmugan, "Computer Aided Tech-

niques for Communications Systems Engineering," IEEE Communications Mag., Vol. 25, pp. 48-54, July 1987.

[14] P.M. Hahn and M.C. Jeruchim, "Developments in the Theory and Application of Importance Sampling," IEEE Trans. Commun., vol. COM-35, pp. 706-714, July 1987.

SOME VISION ALGORITHMS

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1987

Johnson Space Center

Prepared by:	Gordon G. Johnson
Academic Rank:	Professor
University & Department:	University of Houston Department of Mathematics Houston, Texas 77004
NASA/JSC	
Directorate:	Mission Support
Division:	Mission Planning and Analysis
Branch:	Technology and Applications
JSC Colleague:	Robert Savely
Date:	August 14, 1987
Contract Number:	NGT 44-001-800

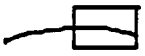
ABSTRACT

The ability to determine the contents of a scene by a machine has obvious applications. There are reasons for seeking hoizontal, vertical and diagonal line segments of varying lengths in order to interpret the content of a scene. Such line segments aid in determining 'edges' that determine the boundaries of objects within the scene. Algorithms for each of the above as well as for data compacting are presented.

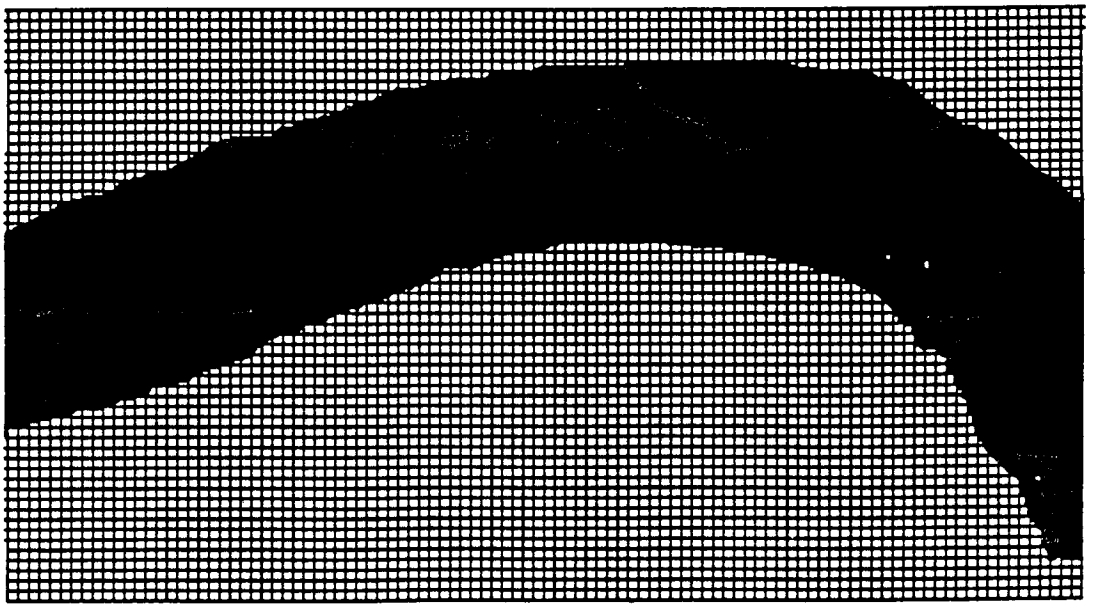
SCENE ANALYSIS

There is, for obvious reasons, an interest in having machines analyze scenes for content. A scene usually contains many objects each having many features. No object can be recognized unless it is already known. This does not mean to imply that the particular object must be known, but rather that the class or classes of objects to which it belongs must be known. The machine then identifies the object as belonging to a class and so indicates. If a significant portion of the scene data set cannot be reconciled with any known data, then the possibility of defining a new object class exists.

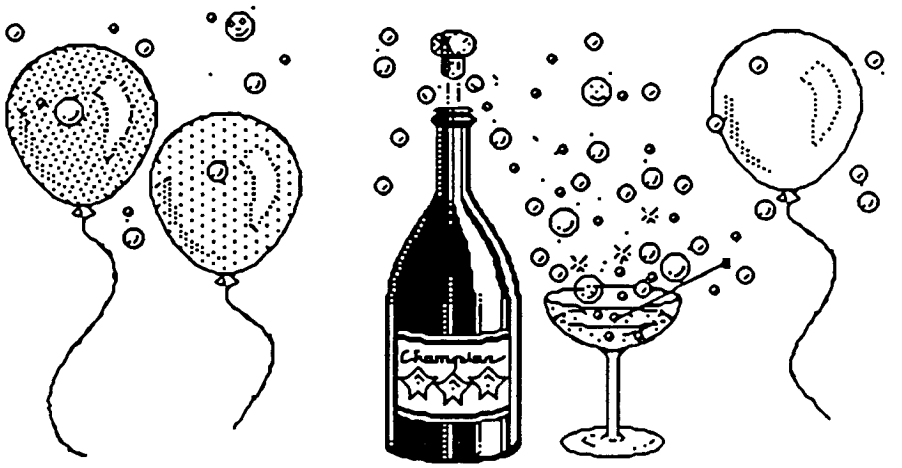
When attempting to process and recognize written letters, a portion of the writing is analyzed at a time, not letter by letter (although this maybe required from time to time when an ambiguity occurs). There is the understanding that the writing, as indicated below, is really studied from a magnified view.



Magnify the above to produce that which is below.

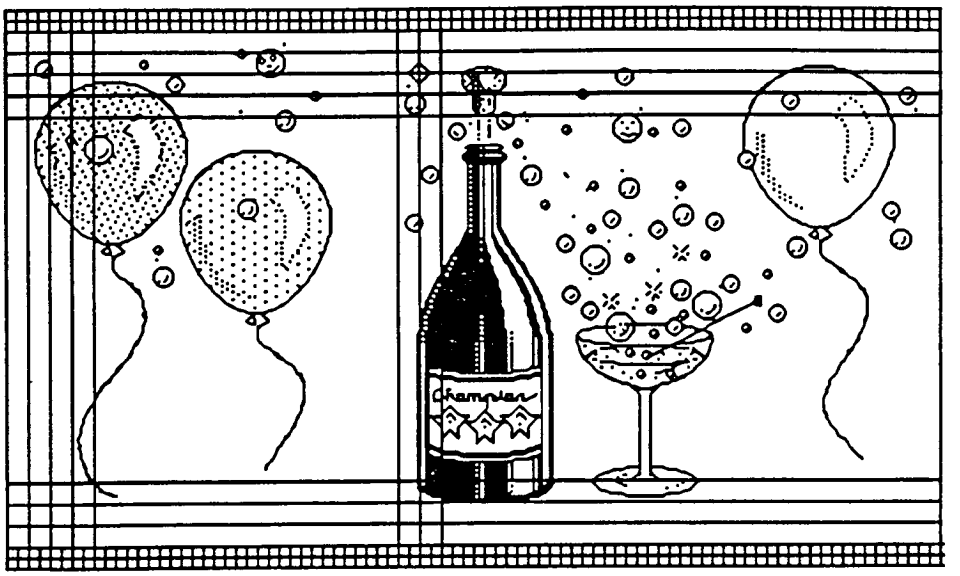


This then is processed by data reduction (explained later), which reduces the data to elementary forms that are then analyzed .



In a complex scene such as the one above, the first task is to identify prominent objects in the scene. This is to be accomplished by first preprocessing of the raw data by data reduction followed by a search for line and edge characteristics. If the nature of the scene is known, then the size of the prominent object or objects normally expected would determine the length of the straight line segments used in the horizontal, vertical and diagonal line searches. On the other hand, if the nature of the scene is unknown, then a standard line length is used, usually one tenth scene height.

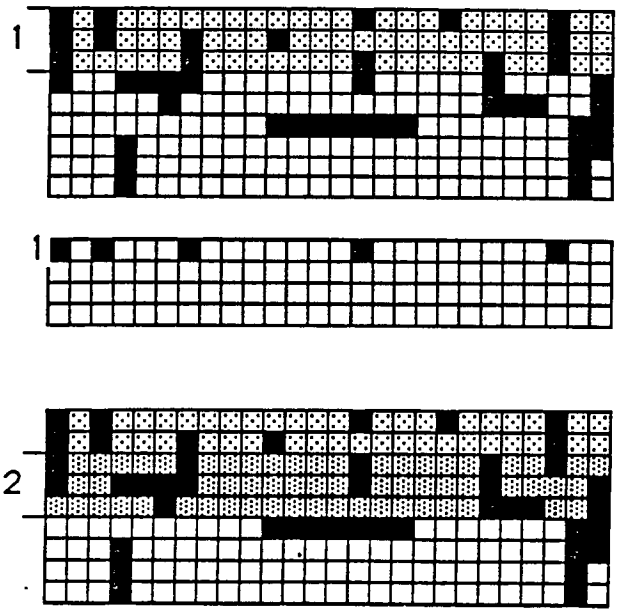
The reduced scene data is first examined for coarse or large figure characteristics formed from line segments. These characteristics allow identification of the classes to which the prominent object or objects in the scene belong.

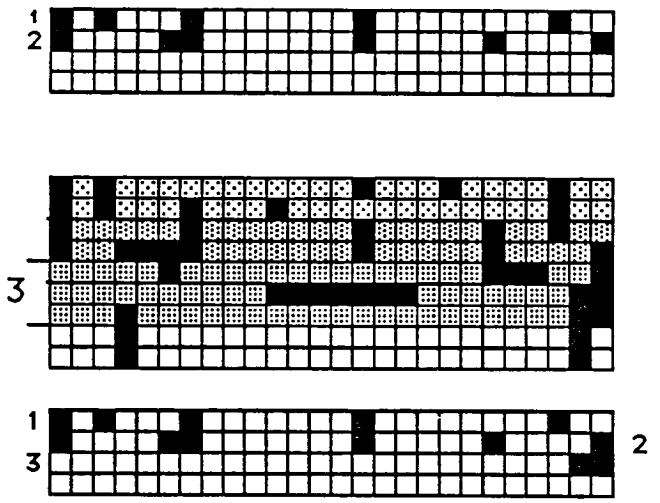


The small object content of an entire scene is expected to be difficult, unless the search is guided by information known about the already identified prominent objects. This information about a particular prominent object would allow the grouping together only thosre smaller objects that could normally be expected to be detected in the region of the particular prominent object. In the event that a significant portion of the data from a region cannot be attributed to any particular object within the restricted grouping, then the analysis is extended to 'out of place' objects.

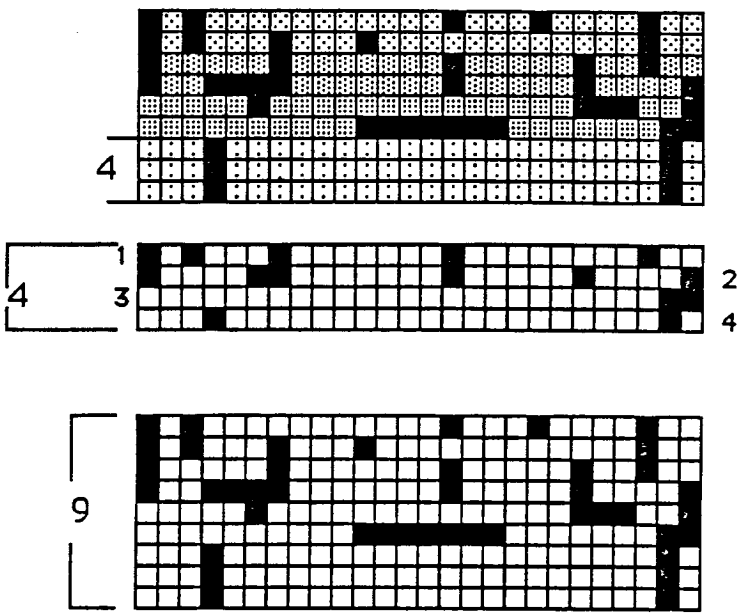
DATA REDUCTION

We now describe how the raw data can be reduced. This reduction reduces the data set by a factor of four. The data is reduced first vertically and then horizontally by considering the top first three rows and determining, in each column of lenght three, if two or more squares are black, in which case a black is saved, else a white. After completing the above for each column, the rows three through five are processed in the same manner. Note that there is a overlap of one row with the first three rows

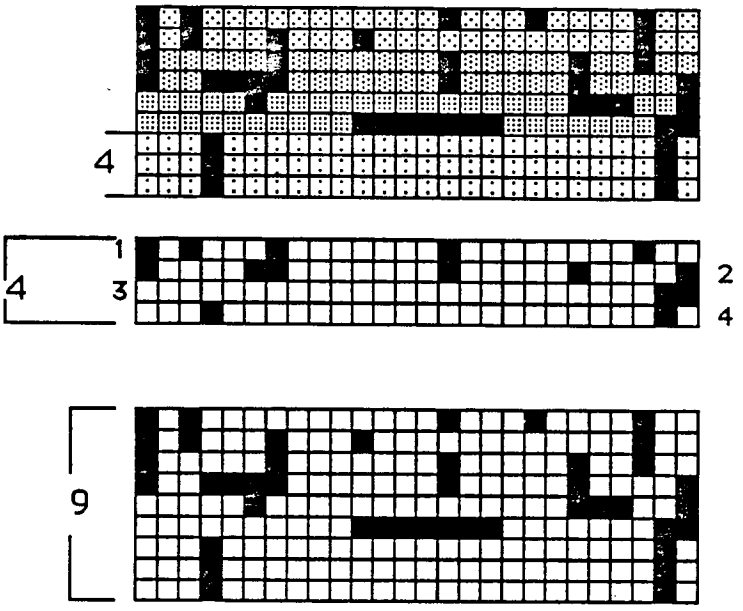




We are now at the last stage in the vertical portion of the reduction.

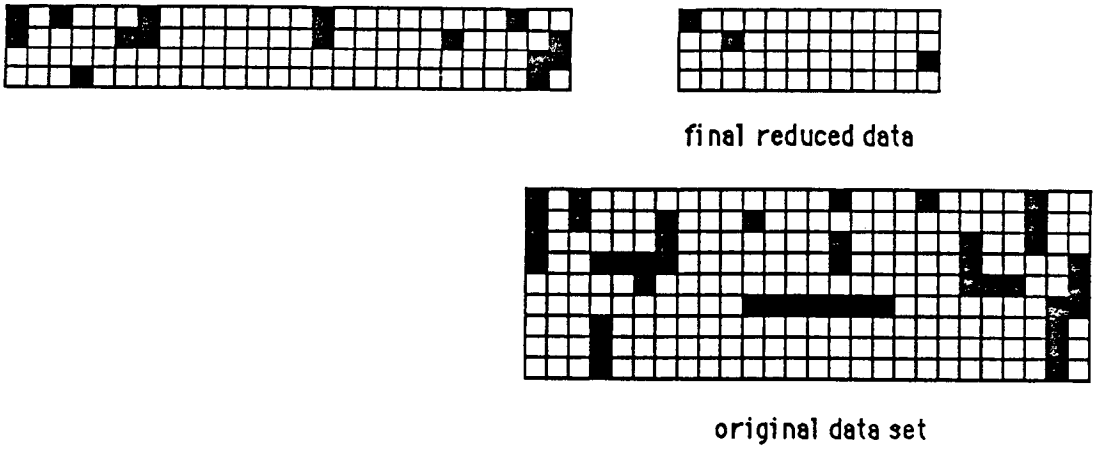


We now are at the last stage in the vertical portion of the reduction.



The last figure is the original data set and the figure immediately above is the data reduced vertically.

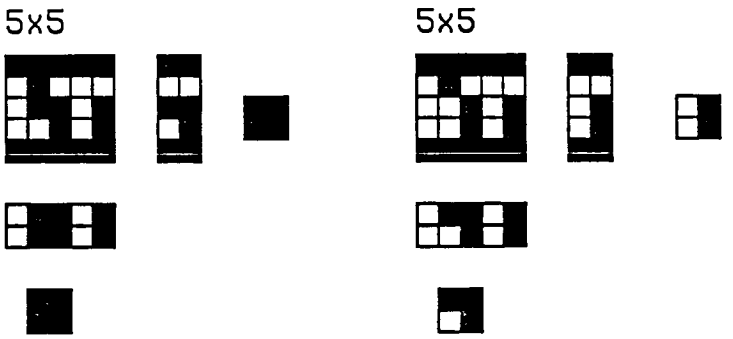
Below we complete the process by the horizontal reduction.



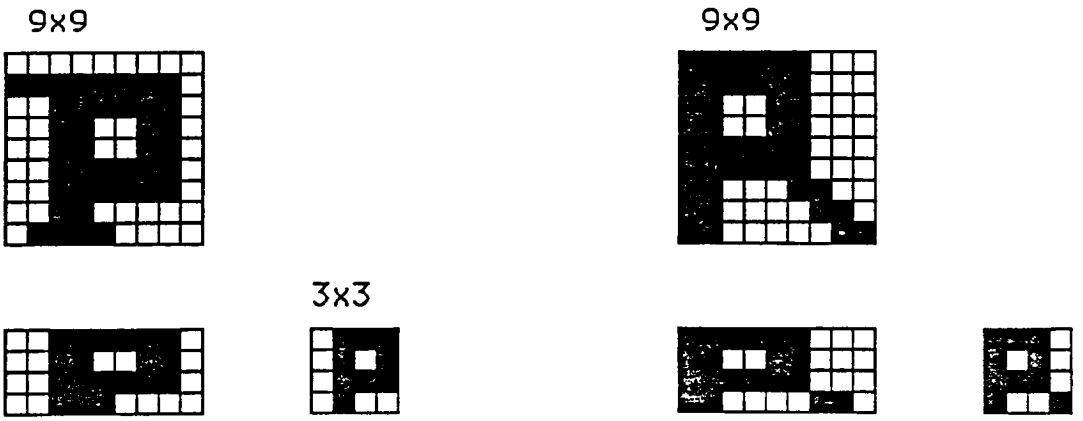
final reduced data

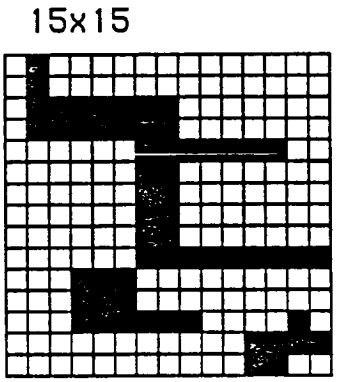
original data set

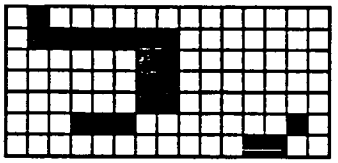
Below are several more examples, where the vertical reduction is followed by the horizontal reduction resulting in the final reduced data set.



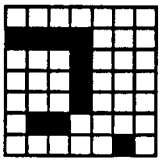
Notice that the order of reduction produces different results in the two examples above.







7x7

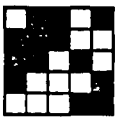


There are other data reduction schemes one of which is in when the first five rows are analyzed looking for three or more blacks in each column, then proceeding to the rows four through eight again looking for three or more in each column. This is continued until the last rows is processed. Notice that this has a two row overlap, as against a one row overlap in the three row case. The three row case results in a reduction by a factor of four while the five row case results in a factor of nine.

The reduction process can be repeated on the reduced set producing a doubly reduced data set.

LINE SEGMENT SEARCH

We now turn to the analysis for determining line segments contained within the reduced data. Data is usually in the form of a square matrix. We create two matrices that we shall use to determine hoizontal, vertical and diagonal line segments contained in the data.



Original or reduced data

$$D = \begin{bmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Matrix for horizontal line segment search of length three.

(Data matrix is to be on the left)

$$H = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

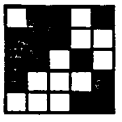
The computation is DxH = HS1, where HS1 is the matrix with the raw horizontal line segment data. We now decide a minimum cutoff value. When this is determined, then the final horizontal line segment matrix is HS.

$$DH = \begin{bmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$HS1 = \begin{bmatrix} 2 & 2 & 2 \\ 3 & 2 & 1 \\ 2 & 2 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

If the cutoff value is three, then we have the matrix below.

$$HS = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$



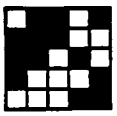
Notice that there is a 1 in the first position in the second row in HS which corresponds to the three black dots in a row in the data grid.

If the cutoff value is two, then we have the matrix below.

$$HS = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

C-4

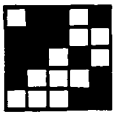
Notice that, when broken into blocks of three, that the three ones in the first row of HS correspond to the first row in the data grid, the two ones in the second row correspond to the three black dots of the second row and so forth.



Matrix for vertical line segment search of length three.
(Data matrix is to be on the right)

$$V = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

The inital vertical line segment computation is VxD =VS1 that contains the raw vertical line segment data. Again a minimum cutoff value is to be decided. Once this minumum is decided, then the final vertical line matrix is VS.



$$VxD = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$VS1 = \begin{bmatrix} 2 & 3 & 2 & 1 & 1 \\ 3 & 2 & 1 & 1 & 1 \\ 2 & 1 & 0 & 2 & 2 \end{bmatrix}$$

If the cutoff value is three, then we have the vertical line matrix below.

$$VS = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

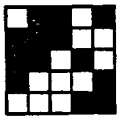
Notice the one in the second position of the first row which corresponds to the three vertical dots in the second column of the data grid, and the one in the first position of the second row which corresponds to the three vertical dots in the first column of the data grid.

$$VS1 = \begin{bmatrix} 2 & 3 & 2 & 1 & 1 \\ 3 & 2 & 1 & 1 & 1 \\ 2 & 1 & 0 & 2 & 2 \end{bmatrix}$$

The first column in VS1 corresponds to the first column in the data grid. Recall that there is an overlap of two between each pair of of the rows in the V matrix.

If the cutoff value is two, then we have the vertical line matrix below.

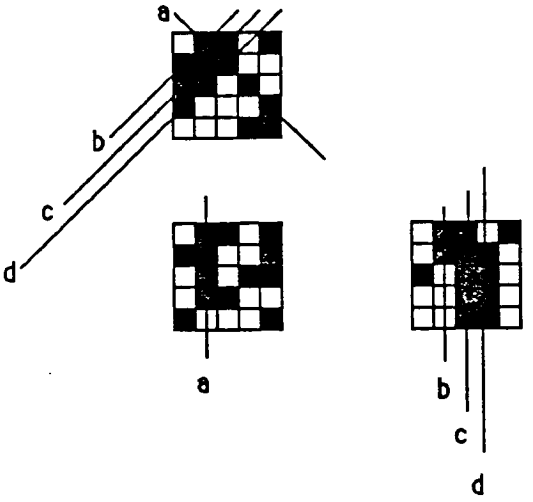
$$VS = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \end{bmatrix}$$

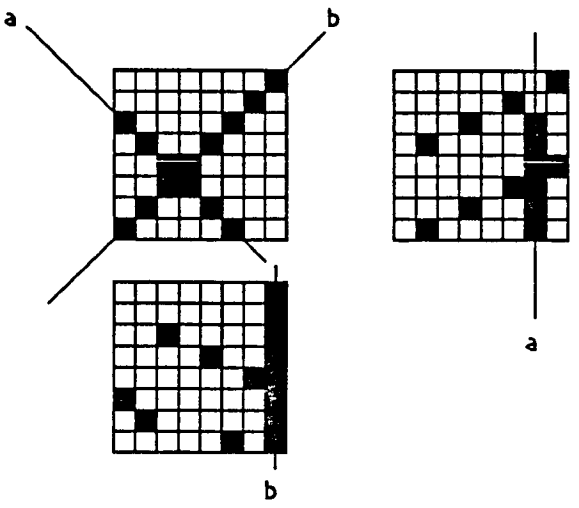


DIAGONAL LINE SEGMENTS

To search for diagonal line segments sloping down from top to the right, we index the second row one to left, third row two to left, etc. and then apply vertical lineseeker.

To search for diagonal line segments sloping from bottom up to left, we index second row one to right, third row two to the right, etc. and then apply vertical lineseeker.





There is may be some difficulties with data from a 'hazy' scene, i.e., a scene where all outlines are blurred such as would be caused by an out of focus camera, very low lighting levels or fog.

References

Frisby, John P., SEEING, Oxford University Press, New York, NY (1980)

Luckiesh, M., VISUAL ILLUSIONS, Dover Pub., New York, NY (1965)

DYNAMICS FORMULATIONS FOR THE REAL-TIME SIMULATION OF CONSTRAINED MOTION

Final Report

NASA/ASEE Summer Faculty Fellowship Program--1987

Johnson Space Center

Prepared by:	Frederick A. Kelly, Ph.D.
Academic Rank:	Assistant Professor
University & Department:	Michigan Technological University Mechanical Engineering - Engineering Mechanics Department Houghton, Michigan 49931
NASA/JSC	
Directorate:	Engineering
Division:	Systems Development and Simulation
Branch:	Teleoperator Systems
JSC Colleague:	Charles R. Price
Date:	August 7, 1987
Contract Number:	NGT 44-001-800

ABSTRACT

The Space Shuttle program has relied heavily on simulation throughout all phases of development and operation. Real-time, man-in-the-loop simulation has served the NASA manned-space-flight program by providing the means to evaluate systems design and integrated systems performance in a simulated flight environment as well as provide a means to train flight crews.

New challenges are presented by the development and operation of a permanently manned Space Station. The assembly of the Space Station, the transferral of payloads and the use of the Space Station manipulator to berth the Orbiter are operations critical to the success of the Space Station. All these operations are examples of constrained motion among the bodies associated with the Orbiter and Space Station system.

Current real-time simulations of the Orbiter and Space Station do not have the capability to model the dynamics of constrained motion. Determining an efficient, yet general, method for constrained motion in multibody systems is one essential key to high-fidelity of these critical operations.

This report described the state-of-the-art of formulating the governing dynamical equations of motion for constrained systems. The uses of the two basic problems in multibody dynamics are discussed. The most efficient formulations of the equations of motion are addressed from the point of view of completeness. The issues surrounding incorporating the constraints into the equations of motion are presented. Finally, an overview of the strengths and weaknesses of the current methods is given, along with some recommendations for further research.

INTRODUCTION

The planned scenarios involving the Shuttle Orbiter, the Space Station and their respective remote manipulators are examples of complex dynamical systems subjected to kinematic constraints, see Figure 1.

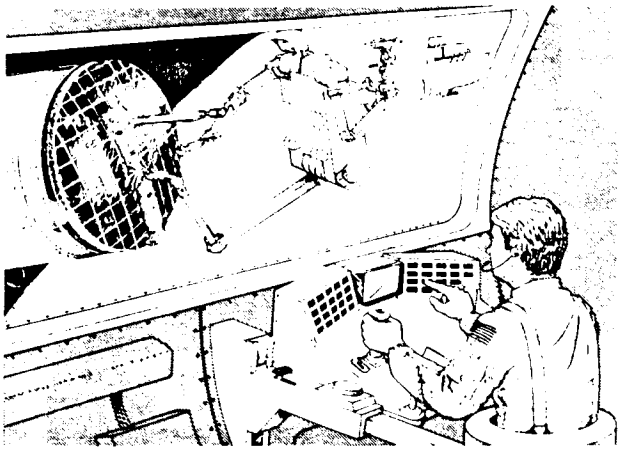


Figure 1 - Constrained motion between TRS and satellite

In order to develop realistic models of these activities, some of the bodies may be considered to be rigid and others flexible. An analysis of the flexibility of the bodies requires a complete knowledge of the system dynamics so that quantitative expressions for the joint forces and moments may be obtained. Having the capacity to formulate completely the equations of motion with the appropriate constraint equations and solve them is essential to any simulation of these systems.

NASA has relied heavily on simulation throughout the Space Shuttle program's development and operation. Real-time, man-in-the-loop simulation has provided the means to evaluate systems design and integrated systems performance in a simulated flight environment and to train flight crews.

Current real-time simulations of the Orbiter and Space Station do not have the capability to model the dynamics of constrained motion. Determining an efficient, yet general, method for simulating constrained motion in multibody systems would enhance the fidelity between the man-in-the-loop simulation and critical on-orbit operations.

This report addresses the efficiency and completeness of current formulations of the dynamics of constrained systems. Efficient formulations are necessary since the goal is to run the simulation in real-time. The formulation should also be complete since flexible bodies may be included.

The balance of this report is divided into three parts with the following part providing a discussion of the equations of motion. The next part discusses incorporating the constraint equations. Concluding remarks are given in the last part.

EQUATIONS OF MOTION

There are two basic problems in multibody dynamics. The first is the problem of determining the motion of a system from a set of applied forces; it is referred to as the forward dynamics problem. The second is the problem of determining a set of forces required to produce a prescribed motion in a system; it is referred to as the inverse dynamics problem.

A solution of the inverse dynamics is essential in the dynamic control of systems and many efficient formulations have been developed to implement real-time dynamic control. A solution of the forward dynamics is essential in the simulation of dynamic systems. Usually the forward dynamics are found by first formulating the inverse dynamics and

then solving for the motion. Recently, several methods for improving the efficiency of the forward dynamics solution have been developed.

The three main approaches towards deriving the dynamic equations of motion for multibody systems with rigid elements have been the Newton-Euler, the Lagrange and the Kane methods. In all these methods efficiency is derived from the structure of the computation. First the linear and angular velocities and accelerations are computed recursively from the reference body to the end of the chain, then the forces and torques are computed recursively from the end to the reference body.

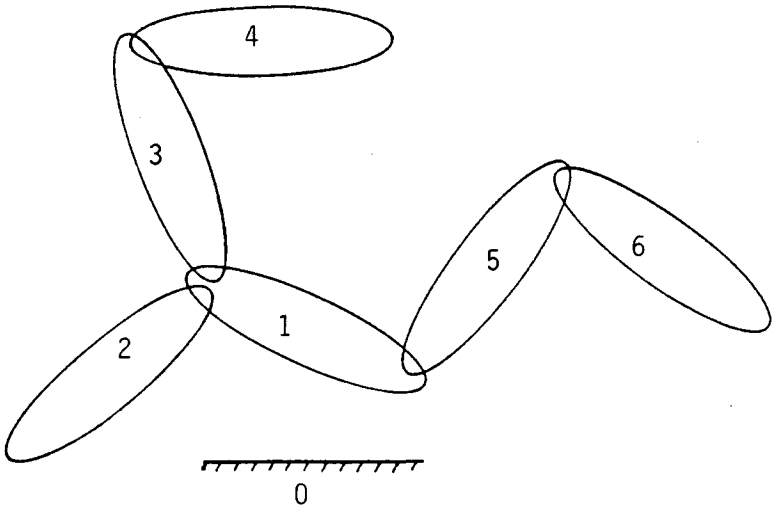


Figure 2 - A general chain system

Consider a multibody system such as shown in Figure 2. An accounting system for the system connectivity may be developed by arbitrarily selecting one of the bodies as a reference body and calling it B_1. Next, number the other bodies of the system arbitrarily. This numbering of the bodies can be used to describe the chain structure or

topology through the "body connection array" as follows: Let L(k), k = 1, ...,N be the array of the adjoining lower body of B_k. The lower body is defined as the body adjoining B_k that is closest to the reference body [1].

The angular velocity of a typical body B_k in the inertial frame, B_o, is readily obtained by the recursion formula

$$\underline{\omega}^k = \underline{\omega}^{L(k)} + {}^{L(k)}\underline{\omega}^k \tag{1}$$

where $\underline{\omega}^k$ and $\underline{\omega}^{L(k)}$ are the angular velocities of B_k and $B_{L(k)}$ relative to the inertial frame and ${}^{L(k)}\underline{\omega}^k$ is the angular velocity of B_k relative to $B_{L(k)}$.

The angular acceleration of B_k in the inertial frame may be obtained by differentiating equation (1) in the inertial frame, leading to the recursion formula

$$\underline{\alpha}_k = \underline{\alpha}^{L(k)} + {}^{L(k)}\underline{\alpha}^k + \underline{\omega}^{L(k)} \times {}^{L(k)}\underline{\omega}^k \tag{2}$$

where $\underline{\alpha}^k$ and $\underline{\alpha}^{L(k)}$ are the angular accelerations of B^k and $B_{L(k)}$ relative to the inertial frame and ${}^{L(k)}\underline{\alpha}^k$ is the angular acceleration of B_k relative to $B_{L(k)}$.

Recursion formula can also be developed for the position, velocity and acceleration of a point on any link [1-4].

A chain system of N rigid bodies will, in general, have 6N or fewer degrees of freedom. This comes as a result of the fact that each rigid body can have up to 6 degrees of freedom. Kinematic constraints between members of the system reduce the number of degrees of freedom through contact between kinematic pairs. Usually, the analysis is simplified by introducing the kinematic constraints due to joints connecting the

bodies of the system at an early stage. The systems's configuration is then said to be completely determined by n generalized coordinates, where n is the number of degrees of freedom of the system. However, it may occur that a contact between bodies may be made or broken as the motion of the system progresses. Figure 3 shows a case where two more bodies of the system in Figure 2 are in contact. In order to distinguish between the two types of constraints the former will be called a permanent constraint and the latter will be called a temporary constraint [5].

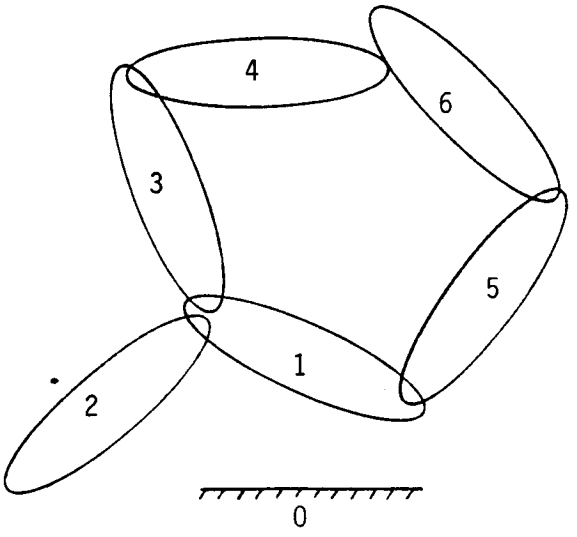


Figure 3 - Temporary constraint between bodies 4 and 6

Many procedures have been developed for both the inverse and forward dynamics of systems with permanent constraints using recursive relationships for efficiency. References [2, 3, 6] represent recently reported research efforts in rigid body inverse dynamics using recursive formulations of the Newton-Euler, Kane and Lagrange methods,

respectively. Recursive formulations of the forward dynamics of rigid body systems with permanent constraints are significantly fewer in number and more intricate in their derivation. References [7-9] represent the currently reported procedures of this type.

Armstrong [7] hypothesized the existence of a linear recursive relationship between the motion of a forces applied to a body and the motion of and forces applied to its neighbors. He defines a set of recursion coefficients for each body and shows how the coefficients for a body may be calculated in terms of those of one of its neighbors. The coefficients are then used to calculate the acceleration of each body. The computational complexity of this method is $O(n)$, but the coefficient of n is quite large. The method is applicable to robots with spherical joints, but a modification is included in Appendix II [7] for revolute joints. This modification increases the computational requirement significantly, although the methods remains $O(n)$.

Walker and Orin [8] describe four methods of solving the forward dynamics problem for systems with revolute or prismatic joints. Three of these methods are based efficient techniques for solving the inverse dynamics problem. The computational complexity of each of these methods is $O(n^3)$, however the coefficients of the higher powers of n are relatively small. The fourth method discussed is based upon a recursive technique for constructing the moment of inertia matrix. The concept of composite center of mass and moment of inertia matrix of a partial set of bodies at the end of the chain is used in this recursion. The fourth method has a computational complexity of $O(n^2)$, however the coefficient of the n^2 term is large.

The basic approach of Walker and Orin's $O(n^3)$ algorithm can be summarized by following Featherstone [9]. The equations of motion can be expressed in the form

$$\underline{\tau} = A(\underline{q})\,\ddot{\underline{q}} + \underline{c}(\underline{q}, \dot{\underline{q}}, \underline{x}) \qquad (3)$$

where $\underline{\tau}$ is the vector of generalized forces acting at the joints; $\underline{q}$, $\dot{\underline{q}}$, $\ddot{\underline{q}}$ are the generalized coordinate vector and its time derivatives; A is the generalized inertia matrix and $\underline{c}$ is the vector of Coriolis, centrifugal and external forces $\underline{x}$.

An algorithm for solving inverse dynamics can be thought of as an implementation of a vector function in the form

$$\underline{\tau} = \underline{d}(\underline{q}, \dot{\underline{q}}, \ddot{\underline{q}}, \underline{x}) \qquad (4)$$

which states $\underline{\tau}$ is obtained from the generalized coordinates, the generalized speeds, the generalized acceleration and the external force acting on the system. It is seen that (cf (3))

$$\underline{c}(\underline{q}, \dot{\underline{q}}, \underline{x}) = \underline{d}(\underline{q}, \dot{\underline{q}}, \underline{0}, \underline{x}) \qquad (5)$$

the calculation of $\underline{c}$ is the special case of calculating the inverse dynamics when the generalized acceleration is zero.

The forward dynamics calculation is conveniently partitioned into three steps: the calculation of $\underline{c}$, the calculation of A, and the solution of the set of simultaneous equations

$$A\ddot{\underline{q}} = \underline{\tau} - \underline{c} \qquad (6)$$

For $\ddot{\underline{q}}$. The computational complexity of the first step is $O(n)$, the second step is $O(n^2)$ and that of the third is $O(n^3)$. This approach is referred to as the composite rigid-body method.

Featherstone [9] describes an extension of Armstrong's method to system with revolute and prismatic joints and uses a spatial notation

consisting of 6X1 spatial vectors, 6X6 spatial transformation matrices and 6X6 spatial inertias. He refers to this method as the articulated-body method.

The articulated-body method starts with the equation

$$\ddot{\underline{q}} = \underline{h}(\underline{q}, \underline{\tau} - \underline{c}) \tag{7}$$

which states that $\ddot{\underline{q}}$ is a function of $\underline{q}$ and $\underline{\tau} - \underline{c}$ without implying the need to calculate a generalized inertia matrix.

The calculation is again conveniently partitioned into three steps: the calculation of $\underline{c}$, the calculation of the inhomogeneous articulated-body inertias for each body starting with the end effector and proceeding sequentially to the reference body, and the generalized accelerations are then calculated for each body starting at the reference body and proceeding sequentially to the end effector. The computational complexity of each step is O(n).

The computational complexity of Armstrong's method is greater than either of the most efficient of Walker and Orin's $O(n^3)$ methods or Featherstone's articulated-body method. The best Walker and Orin method is more efficient than the Featherstone for values of n less than twelve [9]. Either the Armstrong nor the Walker and Orin nor the Featherstone methods provide for the calculation of the constraint forces of the joint. In this sense these algorithms are not complete. The next part describes some methods of incorporating the constraint equations and finding the constraint forces.

INCORPORATING CONSTRAINTS

References [5, 10-15] represent a partial list of recently reported

research efforts on dynamics formulations for constrained motion. The equations of motion for a system of rigid bodies with open chain or tree structure (as in Figure 2) may be written as in equation (6). In systems where additional restrictions have been placed on the motion of the bodies, these constraints can usually be placed in the form

$$B\dot{\underline{q}} = \underline{g} \tag{8}$$

where B is an mxn matrix with m<n and $\underline{g}$ is n vector. Both B and $\underline{g}$ are functions of $\underline{q}$ and time t. Holonomic constraints can always be written in the form of equation (8). Simple or Pfaffian nonhomonomic constraints can also be written in the form of equation (8).

There are two fundamental approaches to solving the equations (6) and (8) for such systems [14]. One can introduce unknown generalized forces of constraint between the constrained bodies by the technique of undetermined multipliers and then solve the equations of motion simultaneously with the constraint equations to determine the generalized constraint forces as well as the kinematic variables. Alternatively, one can use the constraint equations to reduce the dimension of the equations of motion to be solved, eliminating the need to represent or solve for generalized constraint forces. The second approach has the advantage of providing, for solution, a set of equations of minimum dimension. The difficulty lies in its implementation.

In the special case of holonomic constraints, one might integrate equation (8) and solve explicitly for a subset m of the generalized coordinates in terms of the remaining $l \triangleq n - m$. Substitution for these generalized coordinates in equation (6) permits reduction of the set of equations to be solved to l second order equations. In general, this reduction is not easily accomplished.

If equation (8) is nonintegrable (i.e., nonholonomic), then m of the generalized coordinates cannot be solved for in terms of the remaining l. One would like to solve for m of the elements of $\dot{\underline{q}}$, but the result cannot be substituted effectively into the second order equation (6). This obstacle can be avoided by converting equation (6) to first order form [10]. A large obstacle is the potential for the rank r of the matrix B in equation (8) to be less than the row dimension m. These state of affairs could occur in a system changing configuration, resulting in singularities in the numerical solution. Other methods which avoid all singularity problems exist and are described next.

The n generalized speeds, $\dot{\underline{q}}$, may be expressed in terms of l new generalized speeds, $\underline{u}$, by hypothesizing a relationship of the form

$$\dot{\underline{q}} = \beta \underline{u} + \underline{\delta} \tag{9}$$

where β is an nxl matrix and $\underline{\delta}$ is an n vector.

The solution to equation (8) for $\dot{\underline{q}}$ falls into one of two categories. If $\underline{g}$ does not lie in the column space of B (i.e., if $\underline{g}$ cannot be formed by a linear combination of the columns of B), the constraints are inconsistent, the problem is ill-posed and no solution can be found for $\dot{\underline{q}}$. If $\underline{g}$ does lie in the column space of B, at least one solution exists for $\dot{\underline{q}}$. Let Z be an nxl matrix with linearly independent columns that form a basis for the null space of B, that is,

$$BZ = 0 \tag{10}$$

and let Y be an nxm matrix whose columns complete a basis for the vector space R^n. The general soluation to equation (8) is

$$\dot{\underline{q}} = Z\underline{z} - Y\underline{y} \tag{11}$$

where $\underline{y}$ is a vector of m unique scalars such that

$$\underline{g} = B Y \underline{y} \tag{12}$$

and $\underline{z}$ is a vector of l independent quantities describing the constrained system. The vector $\underline{u}$ of generalized speeds for the constrained system may be defined as

$$\underline{u} \triangleq \underline{z} \tag{13}$$

By substituting from equation (13) into (11), one obtains an equation of the form of equation (9) with

$$\beta = Z \tag{14}$$

and

$$\underline{\delta} = -Y\underline{y} \tag{15}$$

The preceding development demonstrates that it is always possible to find an equation of the form of equation (9) for a system subjected to constraints in the form of equation (8). However, it does not give an operational procedure for producing equations (14) and (15). The matrices Z and Y are not unique, and many approaches for constructing them can be envisioned. The problem is equivalent to that of constructing a generalized inverse of B, of which the Moore-Penrose generalized inverse is a particular example. Singular value decomposition [12, 14] and the zero eigenvalue theorem [11], are two methods of calculating this inverse. The remainder of this part will briefly outline two methods using Kane's equations of motion.

Kane's equations of motion [10] may be written in the remarkably simple form

$$\underline{f} + \underline{f}^* = 0 \tag{16}$$

where $\underline{f}$ is the n-vector of generalized active forces and $\underline{f}^*$ is the n vector of generalized inertia forces. The set of equations (6) may be efficiently derived from the set of equations (16) by relatively simple scalar products and recursive kinematic relationships [3, 4].

Kane's equations for constrained systems may be found by solving equation (8) for m of the $\dot{\underline{q}}$ in terms of the remain l, as described before. This leads to a reduced set of equations of motion.

$$\underline{k} + \underline{k}^* = 0 \tag{17}$$

where $\underline{k}$ and $\underline{k}^*$ are l-vectors of reduced generalized active and inertia forces. Equations (17) together with equations (8) then constitute a system of n equations for the n-vector $\dot{\underline{q}}$.

For complex multibody systems the solution of equations (8) for m of the $\dot{\underline{q}}$ may not be convenient or even possible. Also, the generalized constraint forces have been eliminated from the analysis. Although this reduces the dimension of the problem, it also means the analysis is incomplete.

To develop a method which includes the generalized constraint equations and retains the advantageous features of the previous method, let the constraining forces and moments associated with the constraint equations (8) be represented by a series of forces $\underline{f}'_i$ applied at points P_i, i = 1, ...,N', together with a series of couples with moments $\underline{m}'_j$ applied to bodies B_j, j = 1, ...,M'. Assuming that $m \leq 3\ (N' + M')$, the following analysis is applicable [15].

Let the P_i and B_j have specified motions given by $\underline{V}_{Pi}(t)$ and $\underline{\omega}_j(t)$. Let the scalar components of these vectors, relative to a set of basis

vector fixed in the reference body, be collected into arrays $\underline{V}$ and $\underline{\omega}$ whose elements may be identified with the elements of $\underline{g}$ in equation (8). That is,

$$\left.\begin{matrix} V_i \\ \omega_i \end{matrix}\right\} = g_i = \sum_{j=1}^{m} b_{ij}\, \dot{q}_j \tag{18}$$

for $i = 1,\ldots,3N'$ for the V_i and $i = 3N' + 1, \ldots,3(N'+M')$ for the ω_i.

Next, let the components of $\underline{F}'_i$ and $\underline{M}'_j$ in the directions of the components of $\underline{V}_{Pi}$ and $\underline{\omega}_j$ be designated as ϕ_i and μ_j. Then the power, P, produced by the constraining forces and moments is

$$P = \sum_{i=1}^{3N'} \phi_i\, V_i \quad \sum_{j=1}^{3M'} \mu_j\, \omega_j \tag{19}$$

Let $\underline{\lambda}$ be an m-vector whose elements are ϕ_i and μ_j. P may now be written as

$$P = \underline{\lambda}^T \underline{g} = \underline{\lambda}^T B\, \underline{\dot{q}} \tag{20}$$

where the superscript T designates the transpose.

Since neither $\underline{f}'_i$ nor $\underline{m}'_j$ are functions of $\underline{\dot{q}}$, the generalized forces associated with these constraining forces and moments are

$$\underline{f}' = \partial P/\partial \underline{\dot{q}} = B^T \underline{\lambda} \tag{21}$$

where $\underline{f}'$ is the generalized force vector.

In view of the foregoing analysis, let the generalized active forces be separated into two parts consisting of: those developed from constraint forces and moments and those developed from applied forces and moments. Let these be designated $\underline{f}'$ and $\underline{f}$. Then Kane's equations

are

$$\underline{f}' + \underline{f} + \underline{f}^* = \underline{0} \tag{22}$$

Substitute for $\underline{f}'$ in equation (22) using equation (21)

$$B^T \underline{\lambda} + \underline{f} + \underline{f}^* = \underline{0} \tag{23}$$

Taken together with the constraint equations (8) they constitute a set of n+m equations for the n elements of $\dot{\underline{q}}$ and the m elements of $\underline{\lambda}$.

Suppose that an nxl matrix has been found such its columns form a basis for the null space of B. Then by premultiplying equation (23) by β^T we obtain

$$\beta^T \underline{f} + \beta^T \underline{f}^* = 0 \tag{24}$$

since

$$B\beta = \beta^T B^T = 0 \tag{25}$$

Equation (24) together with equations (8) then represent a set of n equations for the n elements of $\dot{\underline{q}}$. Once $\dot{\underline{q}}$ and $\underline{q}$ are determined, $\underline{\lambda}$ may be determined by back substitution into equation (23).

The method may be viewed as a generalization of the psuedo-coordinate method of Kane [10]. The analysis is developed with the system unrestrained. However, a difference is that the constraint forces and moments are formally introduced through a product of undetermined multipliers with coefficients of the constraint equations.

The concept is simple, but the implications extend beyond the task of finding constraint forces and moment components. For example, when

compared with Lagrange's equations with undetermined multipliers, the procedures are similar. However, with Kane's equations the generalized speeds may be employed as the fundamental kinematic variables. Also, holonomic as well as simple nonholonomic constraints may be accommodated directly. Finally, the constraint force and moment components may be either evaluated or eliminated from the analysis. In the latter case the resulting equations are equivalent to the standard form of Kane's equations for constrained systems.

To illustrate these concepts consider Figure 4 which depicts the generalized forces in the n-dimensional space of the generalized speeds. The force triangle is a representation of equation (22). In this context, the generalized forces $\underline{k}$ and $\underline{k}^*$, used in Kane's equations for constrained systems, are seen to be projections of $\underline{f}$ and $\underline{f}^*$ onto the directions β^T, the generalized inverse of B.

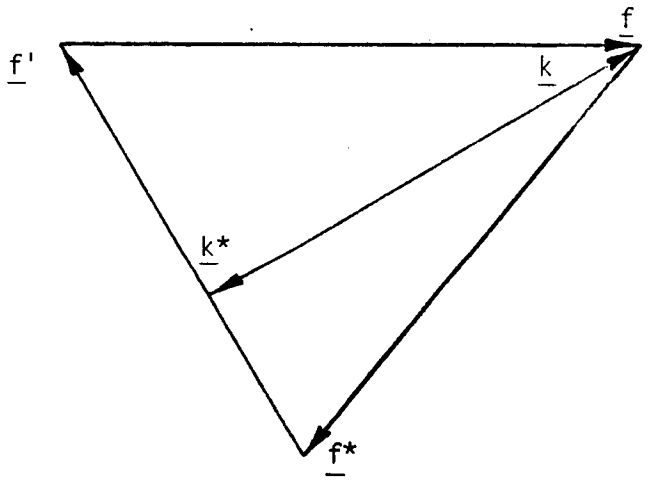


Figure 4 - A geometric interpretation of the generalized forces

CONCLUSIONS

Very efficient procedures have been developed for solving the inverse dynamics problem. Representative solution method using recursive formulatins of the Newton-Euler, Kane and Lagrange methods are described. A generalization of Kane's method [15] has been shown to be capable of efficiently and completely accommodating various constraints applied to a system whereas the others have not. This is not to say that this cannot be done. More research is certainly appropriate to develop parallel techniques using the Newton-Euler and Lagrange methods.

The forward dynamics problem has been more elusive. Although very efficient procedures have been developed for the special cases of permanent constraints of the revolute, prismatic and spherical types, a corresponding generalization has yet to be developed. Indeed, none of the forward dynamic algorithms available can efficiently computer the constraint forces and moments for even permanent constraints. Research is to determine efficient and complete methods of generalizing the solution to the forward dynamics problem, if this is possible.

REFERENCES

[1] Huston, R. L., Passerello, C. E. and Harlow, M. W.: Dynamics of Multirigid-Body Systems. Transactions of the ASME, Journal of Applied Mechanics, Vol. 45, December 1978, pp. 889-894.

[2] Luh, J. Y. S., Walker, M. W. and Paul, R. P. C.: On-Line Computational Scheme for Mechanical Manipulators. Transactions of the ASME, Journal of Dynamics Systems, Measurement, and Control, Vol. 102, June 1980, pp. 69-76.

[3] Huston, R. L. and Kelly, F. A.: The Development of Equations of Motion of Single-Arm Robots. IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-12, No. 3, May/June 1982, pp. 259-265.

[4] Kelly, F. A. and Huston, R. L.: Statics and Dynamics of a Flexible Manipulator. Proceedings of the 5th ASME International Computers in Engineering Conference, Boston, Mass., 1985.

[5] Featherstone, R: The Dynamics of Rigid Body Systems with Multiple Concurrent Contacts. The Third International Symposium on Robotics Research, Gouvieux, France, October 1985, pp. 189-196.

[6] Hollerbach, J. M.: A Recursive Lagrangian Formulation of Manipulator Dynamics and a Comparative Study of Dynamics Formulation Complexity. IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-10, No. 11, November 1980, pp. 730-736.

[7] Armstrong, W. W.: Recursive Solution to the Equations of Motion of an N-Link Manipulator. Proceedings of the Fifth World Congress on Theory of Machines and Mechanisms, Vol. 2, Montreal, 1979, pp. 1343-1346.

[8] Walker, M. W. and Orin, D. E.: Efficient Dynamic Computer Simulation of Robotic Mechanisms. Transactions of the ASME, Journal of Dynamic Systems, Measurement and Control, Vol. 104, September 1982, pp. 205-211.

[9] Featherstone, R.: The Calculation of Robot Dynamics Using Articulated-Body Inertias. The International Journal of Robotics Research, Vol. 2, No. 1, Spring 1983, pp. 13-30.

[10] Kane, T. R. and Wang, C. F.: On the Derivation of Equations of Motion. Journal of the Society for Industrial and Applied Mathematics, Vol. 13, 1965, pp. 487-492.

[11] Kamman, J. W. and Huston, R. L.: Dynamics of Constrained Multibody Systems. Transactions of the ASME, Journal of Applied Mechanics, Vol. 51, December 1984, pp. 899-903.

[12] Mani, N. K., Haug, E. J. and Atkinson, K. E.: Application of Singular Value Decomposition for Analysis of Mechanical System Dynamics. Transactions of the ASME, Journal of Mechanisms, Transmissions, and Automation in Design, Vol. 107, March 1985, pp. 82-87.

[13] Wampler, C., Buffinton, K. and Shu-hui, J.: Formulation of Equations of Motion for Systems Subject to Constraints. Transactions of the ASME, Journal of Applied Mechanics, Vol. 52, June 1985, pp. 465-470.

[14] Singh, R. P. and Likens, P. W.: Singular Value Decomposition for Constrained Dynamical Systems. Transactions of the ASME, Journal of Applied Mechanics, Vol. 52, December 1985, pp. 943-948.

[15] Wang, J. T. and Huston, R. L.: Kane's Equations with Undetermined Multipliers - Application to Constrained Multibody Systems. Transactions of the ASME, Journal of Applied Mechanics, Vol. 54, June 1987, pp. 424-429.

INVENTORY BEHAVIOR AT REMOTE SITES

Final Report

NASA/ASEE Summer Faculty Fellowship Program -- 1987

Johnson Space Center

Prepared by:	William C. Lewis, Jr., Ph.D
Academic Rank:	Research Assistant Professor
University and Department:	University of Washington Mechanical Engineering Dept. Industrial Engineering Prog.
NASA/JSC	
Directorate:	Space and Life Sciences
Division:	Solar System Exploration
JSC Colleague:	Michael B. Duke
Date:	September 14, 1987
Contract Number:	NGT 44-001-800

Abstract

An operations research study is conducted concerning inventory behavior on the Space Station. Historical data from the Space Shuttle is used. The results demonstrate a high logistics burden if Space Shuttle reliability technology were to be applied without modification to Space Station design (which it will not be). Effects of rapid resupply and on board repair capabilities on inventory behavior are investigated.

Introduction

"Adde parvum parvo magnus acervus erit."
"Add little to little, and there will be a great pile."
Ovid, ca. 50 A. D.

Intuition is linear and works best on small systems. Large systems with additive effects are not easy to understand. They produce surprises, like Ovid's pile. Large systems with nonlinear components and additive effects are worse: they produce shocks. These systems cannot be intuitively understood. They must be mathematically understood, or not understood at all.

Inventory systems are nonlinear in important ways, and they are additive as well. This report describes an operations research study of local inventory required to support remote site operations. The study attempts to quantify inventory requirements for generic remote site operations in such a way as to anticipate the order of inventory demands early enough to permit their incorporation into the design process.

As a test case, historical data from Space Shuttle operations is used. This data was very kindly provided by Rockwell International, which deserves much of the credit for the usefulness of these results. In some ways, Space Shuttle data is not appropriate for remote site modeling, in that the Shuttle was designed for ground maintenance after each 1 week mission, whereas the Space Station will never receive ground maintenance, and has a 30 year design life. However, Space Shuttle module reliability was probably increased to the limits of cost effectiveness for 1970 era technology in an attempt to reduce maintenance costs and increase vehicle availability. Also, it is tempting to rely on proven Space Shuttle technology when designing the Space Station. The analysis <u>does not</u> show how Space Station inventory will behave. It does show how Space Station inventory would behave <u>if</u> Space Station technology had the same overall failure rate and repair time distribution as Shuttle technology, which it most

certainly will not.

More significant than the actual numbers is the demonstration that inventory stockouts (that is, exhaustion of inventory for some classes of modules) will always be a factor. If there is a module failure, and no modules of that type are in inventory, there are three courses of action: 1) ignore the failure, 2) repair an already failed module, or 3) get a replacement module by ex tempore resupply. If the failure is in a redundant system, option (1) may be attractive. If not, (2) or (3) will be the options of choice. The results of this analysis strongly suggest that (2) and (3) will both be desirable during Space Station operations, as they are at any remote site operation.

It is discovered on analysis of Space Shuttle operations records that systems comparable to the Space Shuttle require unexpectedly large inventory for long missions. It is further discovered that, even with a fairly large spares inventory, shortages will occur at a significant rate. Makeup of these shortages, either through repair and improvisation, or through special equipment delivery, may be a significant problem in space station and planetary expedition logistics.

The mathematics, the data, the analyses, and the conclusions are described separately to permit replication by the interested reader. A reader not interested in replication or time consuming development of mathematical intuition in this area may skip to the results section.

Mathematical Approach

The techniques used are elementary. Elementary statistical methods are to be preferred over advanced methods on grounds of clarity if they correctly model all significant aspects of a situation. The approach described below is correct in the sense that it provides an estimate of stockout probability distribution parameters which is moderately reliable and can be shown on theoretical grounds to be insensitive to actual item failure distributions, which are not ordinarily known. I make no claims to the method's optimality if actual item failure distributions are known exactly.

The statistical method used to set inventory stock levels and predict number of stockouts is developed below:

1) Independent actors, each performing the same act with an idiosyncratic distribution of times between performance, will as a whole perform the act with a negative exponential distribution of times between

performance. For example, each person in a dialing area makes telephone calls on an individual basis. The time between calls for any individual is a random variable, and the distribution with which this variable is associated will, in general, vary from one person to the next. That is, each person's time between calls will have an associated and unique, or idiosyncratic, distribution. If calls made by all people in the dialing area are observed (by monitoring call arrivals at the telephone exchange), it will be observed that the time between arrival of new calls will behave like a random variable with a negative exponential distribution. The same argument holds true for arrival of customers at a bank, arrival of cars at a gas station, or "arrival" of equipment failures. The density function for a negative exponential distribution is:

f(Lambda,t) := Lambda * exp(-Lambda*t)

"Lambda" is the average arrival rate, in events/unit time. "t" is the time between events. It is obtained by adding up individual arrival rates for all independent actors. The average time between arrival is, of course, 1/Lambda; the variance in time between arrivals is 1/(Lambda^2). Thus, high rates of arrival give a smaller variance in interarrival time than do large rates. That is, the process varies more at low arrival rates than at high.

2) Any process with density function f(Lambda,t) of interarrival time will have a Poisson distributed number of events per unit time. That is, if instead of observing time between events, one counts events per unit time, one will find that the number of events per unit time is Poisson distributed if and only if the time between events has a negative exponential distribution f(x).

The Poisson distribution is as follows:

P(Lambda,N):= exp(-Lambda)*(Lambda^N)/N!

"Lambda" is defined as before (and is indeed identical to the Lambda for f(Lambda,t)). Lambda is also the mean and variance of P(Lambda,N). Note that variance of P(Lambda,N) increases with increased Lambda, whereas that of f(Lambda,t) decreased under the same condition. The higher the arrival rate, the more variance, hence the more spread in observed number of arrivals over several observations. "N" is the number of events. Note that N is a cardinal number, that is a member of the set {0,1,2,...}. N cannot be 3.5, or -2, as these cannot count events. This will cause certain minor problems later, and has the immediate consequence that P(Lambda,N) is not a density function, but rather a distribution. In general, the probability of N failures per unit time will

be given by P(Lambda,N). The above can be re-written recursively as:

$$P(Lambda,0) := \exp(-Lambda)$$

$$P(Lambda,N) := P(Lambda,N-1) * Lambda/N$$

The behavior of the function P(Lambda,N) as expressed recursively is clear. If Lambda>1, it P(Lambda,N) increases until N>Lambda, then decreases. If Lambda<1, P(Lambda,N) always decreases. Note that P(Lambda,N) is always positive and finite, as one would expect a probability to be.

3) In planning for inventory, one might be interested in having just enough inventory to avoid stockout. That is, one wants to have enough in inventory so that one never runs out, and at the same time wants nothing superfluous in inventory so that one does not waste money. This is impossible, as will be shown.

If failures are Poisson distributed, then the probability of having N failures or less is given by the cumulative Poisson distribution,

$$C(Lambda,N) := sum(0,N,P(Lambda,N))$$

Obviously, the probability of covering all failures when there are N items in inventory and Lambda is constant is just C(Lambda,N). By the same token, the probability of stockout must be 1-C(Lambda,N).

The behavior of 1-C(Lambda,N) is therefore of interest. Let us re-write:

$$1-C(Lambda,N) = sum(N+1,infinity,P(Lambda,N))$$

and we have seen above that P(Lambda,N) will be finite (although possibly very small) for all cardinal numbers N.

From this it is apparent that 1-C(Lambda,N) will always be finite, no matter how large N, and that (accordingly) there is always some probability of stockout, no matter how large N, the number of replacement units in inventory.

4) Fortunately, the above is not merely an impossibility proof. It also leads to an algorithm for sizing inventory. If one is satisfied with a fairly small desired probability of stockout per mission, Psd, then, if Lambda per mission is known, one need only specify an N large enough such that:

$$1-C(Lambda,N) < Psd \quad \text{and} \quad 1-C(Lambda,N-1) > Psd$$

N is the smallest number of items which can be kept in store if Psd is to be met. N is uniquely specified by this algorithm, which may be written as N(Psd). The actual probability of stockout is, of course,

Psa:= 1-C(Lambda,N(Psd))

Lambda for the mission is Lambda per unit time multiplied by number of units of time per mission. Note that, in general, Psa < Psd. This is a result of N being a cardinal number, and is the inconvenience mentioned above. The probability of 1 stockout is P(Lambda,N+1), the probability of 2 stockouts is P(Lambda,N+2) and so on. The expected number of stockouts for the entire mission is thus E(s):= sum(n=N,n=infinity,P(Lambda,n)). Ordinarily only a few terms of this sum must be calculated before P(Lambda,n) becomes negligible.

5) One must somehow specify Psd. For inventories containing only one kind of item this is simple. One defines a cost function Ma(N(Psd)) and minimizes it. For inventories containing several noninterchangeable classes of items, each with its own Psaj (j being an index describing the kind), but supporting a single system (such as the Space Station), this procedure is not satisfactory. The cost function is properly a function of system reliability, which is a function of subsystem reliability, which is in turn a function of the number of working modules within each subsystem, which is decreased by each stockout. While the Psaj must obviously meet certain criteria, because these criteria are dependent on the criticality of the subsystem supported by units of kind j, there emerges an additional, system wide, criteria: the total number of stockouts during a mission (TSM). If j is over about 10, one would expect system degradation and a TSM nonlinearly related to the Psaj.

6) By the reasoning in (1), the random variable representing time between stockouts for the inventory as a whole should have a negative exponential distribution. TotalLambda should have a value equal to the sum of the hazard rates over all classes in inventory. The hazard rate for each kind in inventory over the entire mission is the expected value of stockouts the kind, E(sj). That is,

TotalLambda:= sum(all j, E(sj))

7) Given the distribution types, and TotalLambda, one can calculate probability of having N or fewer stockouts per mission by the cumulative Poisson as above:

C(TotalLambda,N)

This can be used to calculate a range of stockouts most likely to be seen. We assume that things which happen less than 5% of the time are often not seen; thus, the smallest N seen would be Nsmall such that

$$C(TotalLambda, Nsmall-1) < 0.05$$
and
$$C(TotalLambda, Nsmall) > 0.05$$

Similarly, the largest N seen would be Nlarge such that

$$C(TotalLambda, Nlarge-1) < 0.95$$
and
$$C(TotalLambda, Nlarge) > 0.95$$

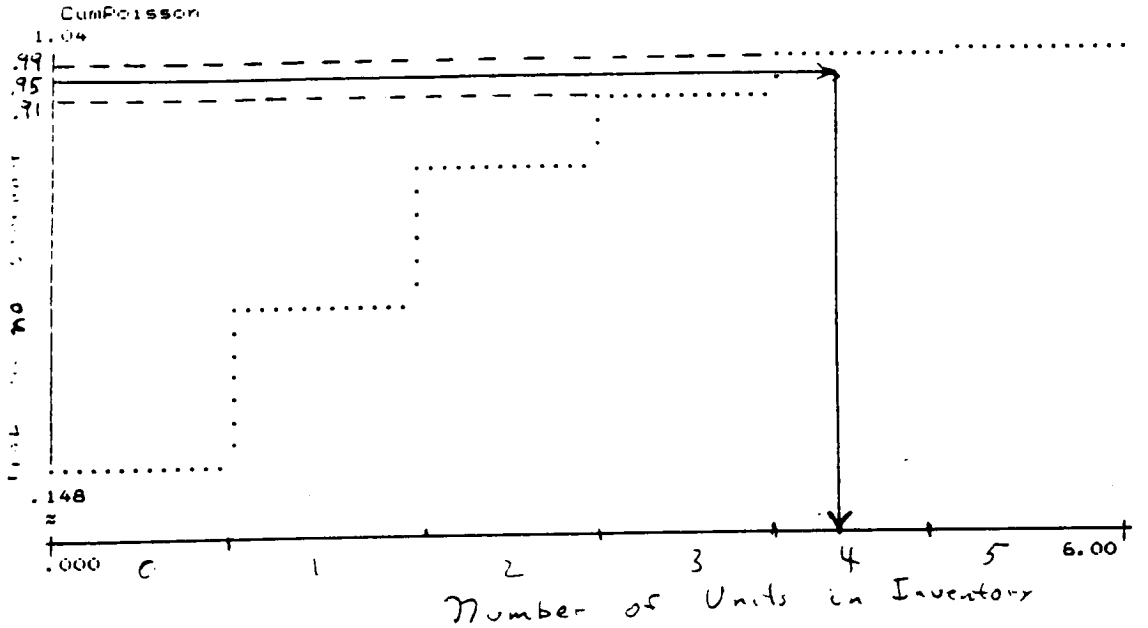
These are approximations, of course, and merely show that one has at least a 5% chance of seeing Nsmall or less and at most a 5% chance of seeing Nlarge or more.

That is a narrative description of the technique. Perhaps it should be illustrated graphically.

The inventory of one kind of unit could be sized by drawing its Cumulative Poisson [C(Lambdaj,N)], and using the curve to translate Psd into minimum N as in Graph A, which looks like a flight of stairs. It is obvious that increasing the number of units in stock from 3 to 4 will reduce Psa to below the 5% minimum specified.

By the same token, given TotalLambda for the completed inventory, stockouts could be determined by drawing its Cumulative Poisson [C(TotalLambda,N)], as in Graph B, and using the curve to translate probabilities of .05 and .95 into numbers of stockouts as shown. Again, it is obvious that the 5% limits of observation lie somewhere around 33 and 55.

Buried in the above analysis is the assumption that stockout hazard rates are constant. In fact, they increase as stock on hand is consumed. In the normal Space Station case, inventory will not be severely depleted during the mission. Perhaps half of it will be used up, and perhaps 1.5% of the inventory classes by module kind will have stockout. One could therefore confidently predict an increase in stockouts during the latter part of a mission. Unfortunately, there seems to be no easy way to solve analytically for stockout distribution as a function of time, and simulation would be necessary to determine this distribution. The distribution given is a valid average distribution for the mission; one would expect the lower limit to be more



C:LIST. Line 1 Col 1 Insert Indent Tab

List of function values.
CumPoisson

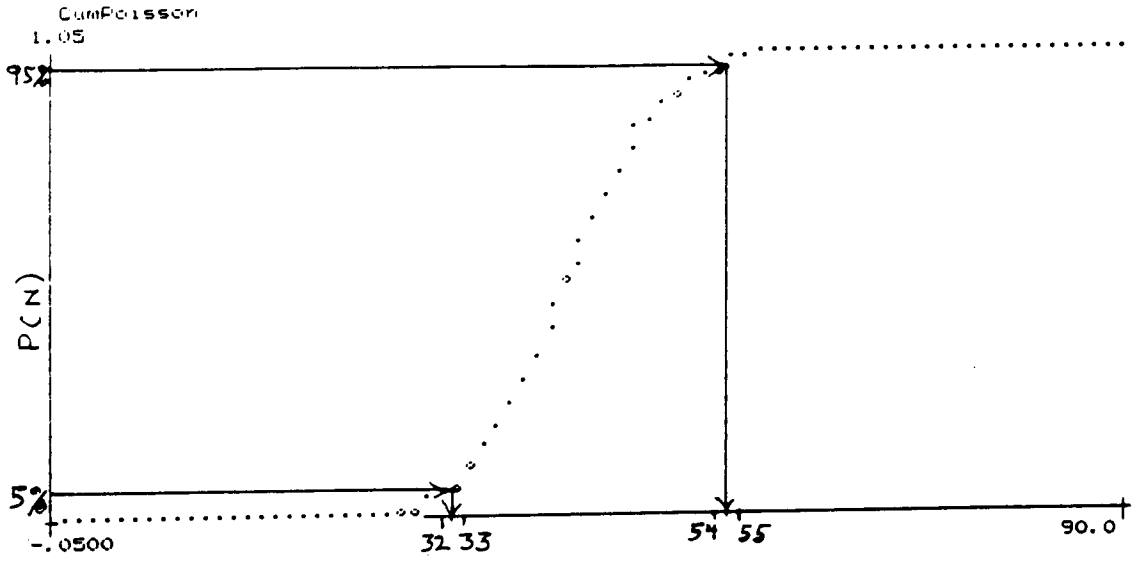
x	CumPoisson(x)	
.00000000	.18824707	
1.0000000	.50261967	
2.0000000	.76512079	
3.0000000	.91124641	← 3 modules ⇒ 8.9% chance of stockout.
4.0000000	.97225386	← 4 modules ⇒ 2.8% chance of stockout.
5.0000000	.99263035	← 5 modules ⇒ .74% chance of stockout.
6.0000000	.99830180	

$\lambda = 1.67$ failures/mission

(6 modules supported, @ 7779 hrs MTDF/Module)

Graph A

$\lambda = 43.4$ stockouts/mission



N Number of Stockout Events

List of function values.
CumPoisson

x	CumPoisson(x)	
30.000000	.020759638	
31.000000	.030820003	
32.000000	.044454650	← 5%
33.000000	.062373498	
34.000000	.085230079	

List of function values.
CumPoisson

x	CumPoisson(x)	
52.000000	.91406814	
53.000000	.93423665	
54.000000	.95043461	← 5%
55.000000	.96320716	
56.000000	.97309883	

Graph B

typical of the mission's early part and the upper limit more typical of the mission's later part. Again, simulation is necessary to if the distribution is to be numerically specified as a function of time.

Data

Two kinds of data can be used. The first set is for initial sizing studies, and is approximate. A uniform MTBR (Mean Time Between Removals) of 8 000 hours, MTTReplace (Mean Time To Replace) of 4 hours, MTTRepair (Mean Time To Repair) of 80 hours, and average commonality of 2 items per type are reasonable starting values.

Historical data on Shuttle subsystem characteristics was provided by Rockwell International. The data set used was the Logistics Master Control File, Report No. 9100, as of 07/13/87. Only QPEI / WUC, MDR, and MTBF data were used. QPEI / WUC (Quantity Per End Item) is the number of the item per individual OV. MDR (Maintenance Demand Rate) is number of unexpected removals per each 1000 hours of OV operation. Therefore, MTBR (Mean Time Between Removals) = 1000/MDR. MTTRepair (Mean Time To Repair) is the actual time on the repair bench required to restore a removed module to fully tested and operational status. RTAT is the number of hours including handling required for module repair. Generally, RTAT>>MTTRepair. Since handling in the sense of transport to repair facilities will not be a major factor at a remote site, RTAT was not used in the analysis. MTTReplace (Mean Time To Replace) was arbitrarily set to 4 hours.

There are 2763 classes of modules in Report No. 9100. Since STOCK2 accepts only 400 module types, some data selection and trimming was necessary.

The first decision was to use only data with MTTR specified. Of 2763 classes, only 1702 (62%) specified MTTR. A t test (see discussion and Table, below) conducted on hazard rate and population supported per class showed no reason to reject the hypothesis of equal means, so the subset was accepted as representative. This subset will be called "MTTR Subset".

The second decision was to select a random subset of MTTR specified data. This involved writing a program (RAND) to generate 400 random numbers sorted in increasing order over a range equal to the cardinality of MTTR Subset. These numbers were treated as indices specifying a second subset of MTTR Subset, called Subset A (which accordingly has 400 classes). A t test (see discussion and Table below) conducted on hazard rate, population supported per class, and mean time to repair showed no reason to reject the hypothesis of equal means, so the subset was accepted as representative. Subset A was

actually used in analyses. It would be desirable to conduct these analyses with a Subset B, C, etc., but time did not permit this.

Statistical characteristics of the data are as follows:

Data Set	Reciprocal Average Hazard Rate	Sample Standard Deviation
Complete	8291 hr	8286 hr
MTTR>0	8264 hr	8264 hr
Not given	8333 hr	8321 hr
Subset A	7231 hr	7231 hr

Data Set	Average MTT Repair	Sample Standard Deviation
Complete	51 hr	253 hr
MTTR>0	82 hr	318 hr
Not given	N/A	N/A
Subset A	93.3 hr	348 hr

Data Set	Average Population Per Class	Sample Standard Deviation
Complete	1.74	2.62
MTTR>0	1.85	2.91
Not given	1.55	2.07
Subset A	1.91	3.04

Note: Total population for any data set can be obtained by multiplying Average Population per Class by Number of Classes.

	Hazard Rate	MTT Repair	Population Per Class
Complete and MTTR Subset	t= 4.72E-4 Nu= 5.9E23	t= 3.41 Nu= 3.0E3	t= 1.27 Nu= 1.2E8
MTTR Subset and Subset A	t= 0.825 Nu= 9.5E21	t= 0.594 Nu= 570	t= 0.360 Nu= 1.5E6

For the Nu values given, the t random variable is distributed normally for all practical purposes. A t<1.282 implies acceptance of the null hypotheses (that data set means are equal) on the 10% level.

The final table, listing t statistics, suggests that the MTTR Subset is very similar to the complete data set except, of course, in MTTRepair. The MTTR Subset and Subset A appear alike for all data types. Accordingly, the behavior of inventory for Subset A should statistically resemble that of the MTTR subset for all data types, and should statistically resemble that of the complete data set for all data types except MTTRepair.

Subset A is stored under file name STOCK2.STK on floppy disk, and will be included with STOCK2. Use STOCK2's export feature to change STOCK2.STK into STOCK2.ASC, which can be imported by a database program as a standard ASCII file with comma as delimiter.

Software and Computer

The software and hardware used are commonly available; those readers wishing to check or extend results may obtain a copy of the software and data on request to the author, if said request is accompanied by a 5 1\4 inch MS-DOS floppy diskette, formatted for a 360k drive. The program is written in Turbo Pascal, and it is advisable to use a database program, such as Borland's Reflex, for data analysis. It should run on any MS-DOS machine. Stock2R.com does not require a mathematics coprocessor, Stock2X.com does. The program is written in Turbo Pascal, by Borland International, and you will need a compiler for this language should you wish to modify and re-compile the program. Please use STK2R or STK2X to install the program if its messages are not properly displayed on your terminal screen. Instructions for the STK2 programs are in document INSTALL.TXT.

A program to generate random numbers uniformly distributed over a specified range is included as RAND and

RANDX. Again, the X suffix means a mathematics co-processor is necessary. Installation is by RND and RNDX. The source file is provided, but requires sort procedures from Borland's Turbo Toolbox to compile.

A random subset of 400 module descriptions from the Space Shuttle's operating history is on the diskette as STOCK2.STK. It is suggested that this information (which includes output data) be imported to STOCK2 by the "typed file" initialization option. Detailed analysis of the results should by choice be attempted with the help of a good database program, as 400 items are too many for the crude display facilities in STOCK2. After alteration by the database program (Reflex by Borland may be a good choice) the data can be re-imported to STOCK2. Use comma as the delimiter.

Although hazard rates scale linearly, confidence limits do not. Accordingly, a program, called LIMITS.com, has been provided to scale the results of STOCK2 up or down. To obtain confidence intervals for a target installation, such as the Space Station, scale up the global hazard rates from Stock2 by:

```
TargetLambda=
  Stock2Lambda * TargetModuleNumber/Stock2ModuleNumber
```

with Stock2Lambda and Stock2ModuleNumber being the values from Stock2, and TargetLambda and TargetModuleNumber being the values of the target installation. For example, if Stock2 gives a repair Lambda of 3.16 for 400 classes of modules with an average of 1.91 modules / class and the Space Station is assumed to have 2578 modules, then the repair hazard rate for the Space Station is:

$$3.16*2578/(400*1.91) = 10.7$$

To obtain Space Station confidence intervals, run LIMITS and specify a hazard rate of 10.7. Use the same scale factor for inventory.

Methods of Analysis

Item demand will change as the population of supported modules ages. If MTBF only changes by a few percent within an inventory period, this can be treated by setting MTBF to the average for the period in question. Calculations involving multiple periods use the appropriate average MTBF for each period. For the purposes of this report, a constant MTBF is the appropriate choice.

A rough feel for this class of inventory problem can be acquired using Borland's Eureka, and interactive equation solver described under "Software and Computer." The algorithm above can be implemented quite directly, and run for simple cases. In particular, it is possible to graphically illustrate the relation between the Poisson and Cumulative Poisson distributions for various values of hazard rate, and to develop one's intuition for stocking levels and stockout probabilities.

In general, one finds that (a) stock levels increase nonlinearly as probability of stockout decreases and that (b) number of modules in inventory divided by number of modules in use tends to decline as population supported increases. The results of such investigations are not reported here, as they are not based on historical data.

When one wishes to use non-aggregated historical data, the sheer mass of computation involved overtaxes Eureka, and a specialized program must be used. This program permits direct entry of equipment characteristics, and executes the mathematical approach described above to determine replacement and stockout hazard rates for each kind of equipment and globally, for the entire inventory. This program, STOCK2, was used to conduct the following preliminary analyses:

1. Inventory behavior under varying uniform stockout probabilities.

Specified probability of stockout for all parts types are set to a uniform value for each run. Inventory size, repair event hazard rate, and 90% confidence interval for number of repair events are calculated.

2. Effect of provision for module repair / cannibalization on inventory behavior and crew size.

Probability of exhaustion is set independently for each part type so as to minimize stockout probability for part types with large repair times, and inventory level is kept constant by increasing stockout probability for part types with small repair times. The net result is to keep inventory size approximately constant while decreasing time required to repair or cannibalize failed modules. After establishing a set of probabilities at average specified probability of stockout of 0.1, probabilities were linearly transformed to obtain other average specified stockout probabilities, and the effect of this on repair time and inventory stock level were recorded.

3. Effect of mission length on inventory behavior.

Longer and shorter mission times are considered with fixed average specified stockout probability. The

shortest time is the length of a Space Shuttle mission; the longest is about that of a round trip Mars mission. The behavior of inventory at an average specified stockout probability of 0.01 is considered.

4. Effect of commonality on inventory behavior.
Greater commonality was simulated by multiplying population supported in Subset A by 5 and by 10, then scaling all results up to the same Space Station number of modules.

There was not sufficient time to conduct these analyses thoroughly. It is likely that only the stronger effects were discovered, and that many weak effects remain as yet unknown.

Results

General:

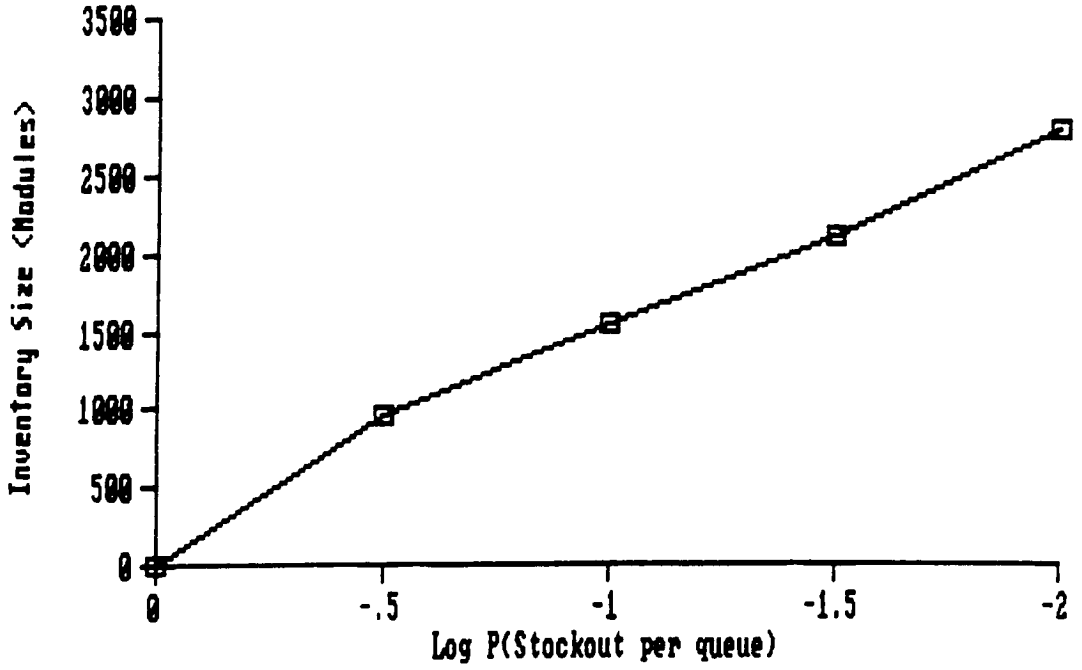
Space Station is assumed to have 2578 ORUs at 1.91 ORU/Class, for a total of 1350 classes.

On at most 5% of the missions values will exceed the Upper Limit; on at least 5% of the missions values will be below the Lower Limit; the two limits thus form an approximate 90% confidence range. An exact 90% confidence range cannot be specified for events, as they cannot be subdivided.

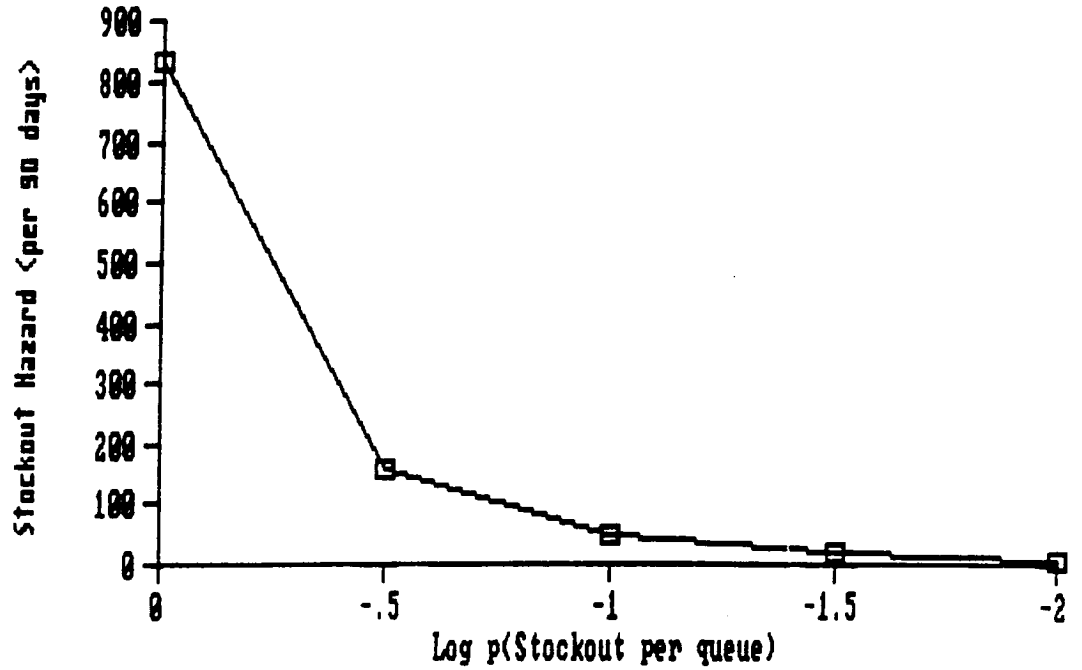
1. Inventory behavior under varying uniform stockout probabilities.

Specified Probability of Stockout	Inventory Size	Stockout Hazard Rate	Number of Stockouts Lower Limit	Number of Stockouts Upper Limit
1.00	0	853	806	901
0.316	967	155	135	176
0.100	1550	51	40	63
0.0316	2110	18	12	26
0.0100	2800	5.2	2	9

Graph C: No Repair
Log P(Stockout) vs. Inventory Size



Graph D: No Repair
Log P(Stockout) vs. Stockout Hazard

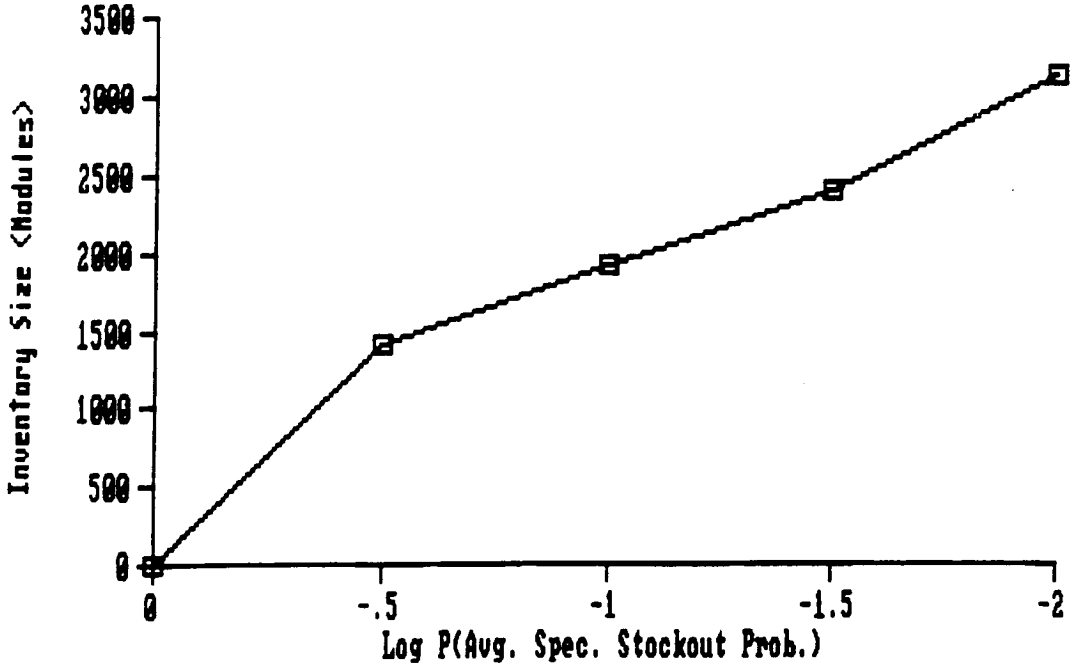


2. Effect of provision for module repair / cannibalization on inventory behavior and crewsize.

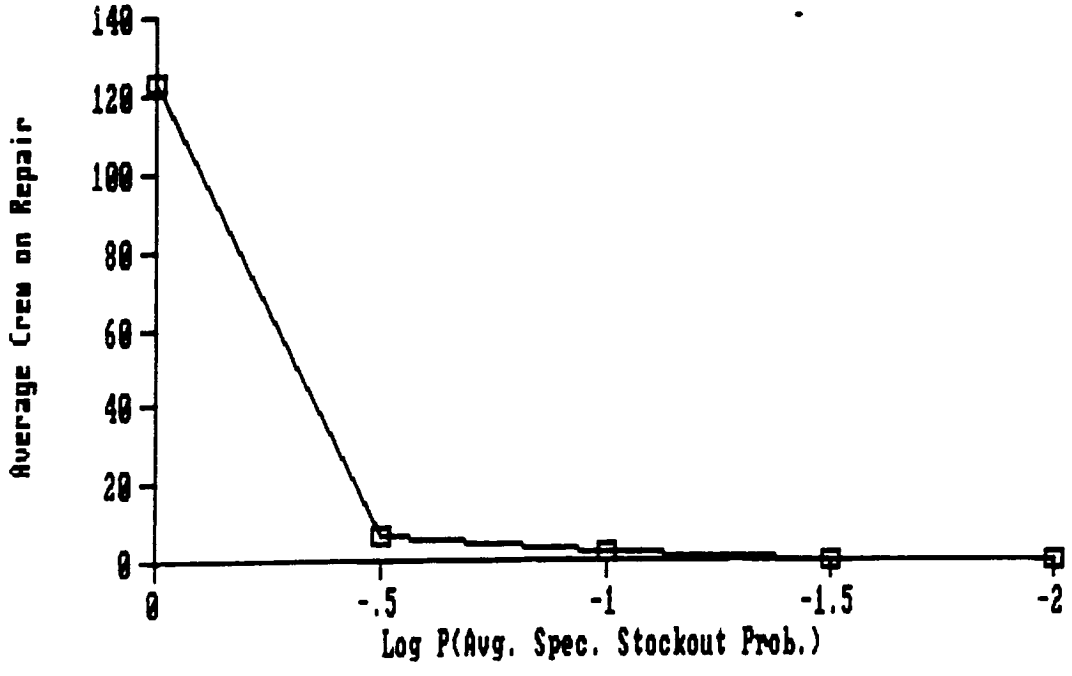
Specified Probability of Stockout	Inventory Size	Stockout Hazard Rate	Number of Stockouts Lower Limit	Number of Stockouts Upper Limit
1.00	0	853	806	901
0.316	1400	102	85	119
0.100	1910	46	35	57
0.0316	2390	18	11	25
0.0100	3150	4.0	1	8

Specified Probability of Stockout	Average Crew on Repair	Repair Hours Hazard Rate	Number of Repair Hours Lower Limit	Number of Repair Hours Upper Limit
1.00	123	88200	87700	88700
0.316	5.9	4220	4110	4320
0.100	1.6	1150	1100	1210
0.0316	0.52	421	388	455
0.0100	0.17	121	103	139

Graph E: Repair
Log P(Stockout) vs. Inventory Size



Graph F: Repair
Log P(Stockout) vs. Crew on Repair



3. Effect of mission length on inventory behavior.

These calculations use data optimized for repair, as given in STOCK3.STK. Average specified probability of stockout is slightly less than 0.1.

Mission Length (Days)	Inventory Size	Stockout Hazard Rate	Number of Stockouts Lower Limit	Number of Stockouts Upper Limit
7	347	10.6	6	16
14	542	15.7	9	22
90	1930	44.0	33	55
270	2560	72.0	58	86
900	12100	97.5	82	114

Mission Length (Days)	Average Crew on Repair	Repair Hours Hazard Rate	Number of Repair Hours Lower Limit	Number of Repair Hours Upper Limit
7	9.5	530	492	568
14	5.9	658	617	701
90	1.6	1110	1060	1170
270	0.9	1910	1830	1980
900	0.4	2660	2580	2750

4. Effect of commonality on inventory behavior.

Commonality	Inventory Size	Replacement Event Hazard Rate	Stockout Event Hazard Rate
1.91	1930	855	44
9.55	1350	852	15
19.1	1270	852	13

Discussion

As can be seen, there is significant intuitive content to the mathematical treatment:

1) Hazard rates are additive.

2) Inventory levels are nonlinear with protection. High protection from stockouts requires a big inventory. Perfect protection requires an infinite inventory.

3) Moderate protection can be accomplished with a much smaller inventory.

4) If there are a large number of separate inventories, the hazards from "moderate protection" add up to significant numbers of observed stockouts.

Mathematical technique, in this case, becomes most interesting when applied to the interpretation of historical data. The intuitive content requires scaling; are we talking about an inventory of 100 or 1,000,000? Will we run out of stock for 1 or 100 items? Historical data for the Space Station is nonexistent; it has not been built. Let us suppose, however, that the Space Station was to be built using the same general order of reliability as the Space Shuttle. This is unlikely, but makes an interesting test of the mathematics, and gives us some idea of the changes in Shuttle era technology required to make a logistically tractable Space Station.

Inventory behavior under varying uniform stockout probabilities is as expected. Stockout hazard rate decreases nonlinearly as specified probability of exhaustion increases. Inventory size increases exponentially with specified probability of exhaustion. In practice, an inventory size of 2800 supporting 2578 running might be chosen to ensure that there will usually be fewer than 9 uncorrected failures per 90 day mission. This is about one item in inventory for each item running.

The effect of provision for module repair and cannibalization is to trade inventory volume for labor. If one is willing to devote the labor of 1.6 people in orbit, inventory can be reduced by 886 modules, from 2800 modules to 1914 modules, and number of uncorrected failures reduced from a maximum of 9 to a maximum of 0. This requires a distortion of inventory policy, however, in that one must stock more of the high repair time modules and fewer of the low repair time modules to obtain the low average repair time which makes on orbit repairs feasible. If only 0.52 people work on repairs, inventory size is reduced only by 411 modules, and assigning 0.17 people actually increases inventory by 351 modules.

Increased mission length, or time between resupply, tends to increase inventory. Accordingly, a doubling of the Space Station's mission from 90 to 180 days would be

expected to increase its required inventory. Unfortunately, not enough points could be taken in the available time to quantify this effect reliably. The algorithm would have to be rewritten so as to keep stockouts/unit time constant (assuming repair) or to keep total stockouts constant (assuming no repair or on demand resupply). However, it is clear that increasing mission length while keeping specified probability of stockout constant will produce significant increases in inventory levels. The 900 day mission requires 4.7 items in inventory for every item running.

The mission length data provides a reality check of the data and mathematics. The predicted removal hazard rate for the Shuttle is 136 removals/mission (90% confidence interval of 117 to 155), whereas Rockwell's data indicate that there are actually 224.5 removals/mission during ground reconditioning. Since not all of the actual removals are for unexpected faults, the prediction seems reasonable.

Specified probability of stockout (SPS) was kept constant for all mission lengths. The effect of SPS is to set a global upper limit on number of stockouts. During the short mission, many of the more reliable modules will not fail, so that the actual probability of exhaustion for that module time will be much less than the SPS for that type, and the global limit will not be approached. As the mission grows longer, the difference between SPS and actual probability of exhaustion diminishes, and the global upper limit is more closely approached. For example, the results listed show a global stockout hazard rate increasing from 10.6 to 97.5. The interesting side effect is that the increase is not linear. That is, the number of stockouts per day becomes less as the mission gets longer, and fewer people are required to repair failures. This suggests that failure repair may be a more attractive option as missions grow longer. Alternatively, by devoting more crew to repair activities, inventories shown might be substantially reduced.

The effect of commonality on inventory behavior is to decrease inventory. While only three points were taken, they suggest that inventory size can be reduced by a maximum of about a half by increasing commonality. It is not clear to what extent commonality can be increased.

Operationally, the above shows conclusively that the Space Station (and other remote site bases) cannot be built under the same design philosophy as the Space Shuttle. The Space Station has a 30 year design life, and does not visit a refitting base. It <u>is</u> the refitting base. Accordingly, it must be kept operational on a continuing basis.

Achieving this will clearly require significant attention to module reliability, repairability, and on demand resupply. Modules must be more reliable to reduce inventory requirements; they must be repairable to minimize the effect of the inevitable stockouts; and rapid resupply of nonrepairable modules must be available on a routine basis. This is not a matter of opinion or engineering judgement. It is implicit in the mathematics.

Learning to do this is a subgoal of the Space Station project, which is (in large part) intended to provide experience in remote site base operation. Proper attention to support factors of the same general type as those considered in this report will result in a Space Station which can be supported, and which will require less ground support as it matures. The lessons learned in Space Station operation will later be applied to design of a Lunar Base, and lessons from the Lunar Base, in turn, applied to a Manned Mars Mission. In this way, logistics problems will be solved incrementally, in a timely and effective manner, and the manned exploration of the Solar System will continue at a steady pace, as foreseen in the Report of the President's Commission on Space and in the Ride report.

Further Work

First, analysis (3) should be carried out more rigorously, providing some convenient way of predicting behavior inventory levels for long missions when those for short missions are known.

The repair hour hazards developed above are really averages. Repair hour hazard is a function of time; it will be higher towards the end of the mission, when inventory is depleted, then it will be at mission start. The only feasible ways of determining the repair hour hazard distribution appears to be simulation. In extreme cases, the repair activity might absorb all available crew labor towards mission end. It would be useful to predict and avoid that.

The existence of high failure rate parts should be investigated further. Preliminary examination suggests that there are surprisingly few such parts in the STOCK2.STK data set, and that reliability improvement would require significant MTBF increase over many part types. For that matter, the effect of increasing MTBF is not properly considered above. It will, obviously, decrease inventory size, but will probably do so nonlinearly. It would be interesting to attempt to characterize the effect in the same way that commonality has been considered above, so as to determine the probable contribution of increased MTBF to reducing the logistics

problem.

STOCK2 could be extended to accept more than 400 data classes. It would be interesting to process the entire data set and compare the results with those of Subset A.

Application of STOCK2 to data sets from other equipment, such as planetary probes and modern commercial aircraft, would provide some badly needed perspective on inventory policy. Data from Antarctic bases, extended Oceanographic expeditions, and industrial plants in countries without a local industrial infrastructure might provide a look at successful inventory policy. Application of STOCK2 to the Hubble Space Telescope might prove particularly interesting.

Failure while in inventory has not been explicitly treated, as it is believed to be of negligible importance. If it is important, some fraction of the modules in inventory must be added to the population of modules supported. For example, if MTBF in inventory is twice that in service, than half of the modules in inventory must be counted as in service. This operation produces a convergent series, and one would expect the convergence to be rapid. Note also that if some of the modules are in active standby, possibly with automatic switchover, it may be desirable to count them as inventory rather than as running. The effect is to reduce the number of "running" modules to the minimum required for station operation. The risk, of course, is that one may find the inventory exhausted with only one "necessary" module in a redundant system running and half the mission to go. Obviously, some difficult and real decisions about what is "necessary" are called for before any such reduction can be considered.

Specified probability of stockout is the only independent variable in the current algorithm. It would be convenient for some kinds of analyses to modify it so as to accept number in stock as the independent variable, and in others to modify it so as to accept number of stockouts as the independent variable.

Optimization of the inventory under the following conditions would be of interest, and has not been addressed above:

* Partial resupply (perhaps by small ELV). Total resupply is treated above as a shortened mission.
* Constrained availability of Criticality 1 modules.
* Repairability of only a subset of module types.

Characteristics of the inventory of replacement components used in repair have not been considered.

In addition to the above, there are related questions concerning ground based logistics, which must have the modules needed for restock available when needed, to include rapid resupply missions. There is also the question of proper techniques, tools, inventory, and so on, used for on orbit module repair. The scope of these questions is beyond what any one person could accomplish, and are suggested as tasks for an organization.

Acknowledgements

The North American Rockwell company was extremely helpful in providing recent shuttle data.

Many people at JSC provided assistance in obtaining data, and listened to earlier versions of this work with considerable patience. Roy Glanville, who provided the data for the first rough calculations, is especially to be thanked.

The Borland International Company, which made most of the support software which permitted this project to be carried as far as it has been during about 2 month's work, is also to be congratulated on the power of their products, as is Symantec, who made the data base and text editing program used to enter and partially analyze the data, and to write this report.

Bibliography

Data used is Rockwell International's Logistic Master Control File Baseline Report, also called Report 9100, dated 07/13/87. Permission to see the original data may be requested from the logistics branch of Rockwell's plant in Downey, CA.

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA CR 172009		

4. Title and Subtitle	5. Report Date
NASA/ASEE Summer Faculty Fellowship Program--1987 Volume 1	November 1987
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
William B. Jones, Jr. and Stanley H. Goldstein	
	10. Work Unit No.
9. Performing Organization Name and Address	11. Contract or Grant No.
Texas A&M University and The University of Houston--University Park	NGT-44-001-800
	13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address	Contractor Report
National Aeronautics and Space Administration Washington, D.C. 20546	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

The 1987 Johnson space Center (JSC) National Aeronautics and Space Administration (NASA)/American Society for Engineering Education (ASEE) Summer Faculty Fellowship program was conducted by Texas A&M University and JSC. The 10-week program was operated under the auspices of the ASEE. The program at JSC, as well as the programs at other NASA Centers, was funded by the Office of University Affairs, NASA Headquarters, Washington, D.C. The basic objectives of the program, which began in 1965 at JSC and in 1964 nationally, are (a) to further the professional knowledge of qualified engineering and science faculty members; (b) to stimulate an exchange of ideas between participants and NASA; (c) to enrich and refresh the research and teaching activities of participants' institutions; and (d) to contribute to the research objectives of the NASA Centers.

Each faculty fellow spent 10 weeks at JSC engaged in a research project commensurate with his/her interests and background and worked in collaboration with a NASA/JSC colleague. This document is a compilation of the final reports on the research projects done by the faculty fellows during the summer of 1987. Volume 1 contains sections 1 through 18, and volume 2 contains sections 19 through 35.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
	Unclassified-Unlimited

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of pages	22. Price*
Unclassified	Unclassified	326	NTIS

*For sale by the National Technical Information Service, Springfield, Virginia 22161